

## Article

# Prediction of Short-Term Winter Photovoltaic Power Generation Output of Henan Province Using Genetic Algorithm–Backpropagation Neural Network

Dawei Xia <sup>1</sup>, Ling Li <sup>1</sup>, Buting Zhang <sup>1</sup>, Min Li <sup>1,\*</sup>, Can Wang <sup>2</sup>, Zhijie Gong <sup>2</sup>, Abdulmajid Abdullahi Shagali <sup>2</sup> , Long Jiang <sup>2</sup> and Song Hu <sup>2</sup> 

<sup>1</sup> State Grid Henan Electric Power Research Institute, Zhengzhou 450052, China; 13526857317@139.com (D.X.); ll\_hust@163.com (L.L.); zhangbuting@163.com (B.Z.)

<sup>2</sup> State Key Laboratory of Coal Combustion, School of Energy and Power Engineering, Huazhong University of Science and Technology, Wuhan 430074, China; m202271281@hust.edu.cn (C.W.); gongzhijie1124@163.com (Z.G.); aashagali@163.com (A.A.S.); jianglong@hust.edu.cn (L.J.); hssh30@163.com (S.H.)

\* Correspondence: 13673368530@139.com

**Abstract:** In the low-carbon era, photovoltaic power generation has emerged as a pivotal focal point. The inherent volatility of photovoltaic power generation poses a substantial challenge to the stability of the power grid, making accurate prediction imperative. Based on the integration of a backpropagation (BP) neural network and a genetic algorithm (GA), a prediction model was developed that contained two sub-models: no-rain and no-snow scenarios, and rain and snow scenarios. Through correlation analysis, the primary meteorological factors were identified which were subsequently utilized as inputs alongside historical power generation data. In the sub-model dedicated to rain and snow scenarios, variables such as rainfall and snowfall amounts were incorporated as additional input parameters. The hourly photovoltaic power generation output was served as the model's output. The results indicated that the proposed model effectively ensured accurate forecasts. During no-rain and no-snow weather conditions, the prediction error metrics showcased superior performance: the mean absolute percentage error (*MAPE*) consistently remained below 13%, meeting the stringent requirement of the power grid's tolerance level below 20%. Moreover, the normalized root mean square error (*NRMSE*) ranged between 6% and 9%, while the coefficient of determination ( $R^2$ ) exceeded 0.9. These underscored the remarkable prediction accuracy achieved by the model. Under rainy and snowy weather conditions, although *MAPE* slightly increased to the range of 14% to 20% compared to that of scenarios without rain and snow, it still adhered to the stringent requirement. *NRMSE* varied between 4.5% and 8%, and  $R^2$  remained consistently above 0.9, indicative of satisfactory model performance even in adverse weather conditions. The successful application of the proposed model in predicting hourly photovoltaic power generation output during winter in Henan Province bears significant practical implications for the advancement and integration of renewable energy technologies.

**Keywords:** photovoltaic power generation; rain and snow weather; genetic algorithm; back propagation; prediction accuracy



**Citation:** Xia, D.; Li, L.; Zhang, B.; Li, M.; Wang, C.; Gong, Z.; Shagali, A.A.; Jiang, L.; Hu, S. Prediction of Short-Term Winter Photovoltaic Power Generation Output of Henan Province Using Genetic Algorithm–Backpropagation Neural Network. *Processes* **2024**, *12*, 1516. <https://doi.org/10.3390/pr12071516>

Academic Editor: Masoud Soroush

Received: 4 June 2024

Revised: 10 July 2024

Accepted: 13 July 2024

Published: 19 July 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Amidst rapid economic and social progress, the global community grapples with the dual challenges of energy scarcity and environmental degradation. Countries around the globe are focusing on achieving sustainable development with reduced carbon footprints and actively pursuing renewable energy alternatives. According to data from the International Energy Agency (IEA) [1], as of 2022, the cumulative capacity of installed photovoltaic systems globally had soared to 1185 GW. Notably, photovoltaic power generation

contributed to a reduction of approximately 1.4 billion tons of carbon dioxide emissions, representing a remarkable 30% increase compared to 2021. This reduction accounted for 10% of carbon emissions from the power and heating sector. In the broader context of the global shift towards low-carbon energy, photovoltaic technology has emerged as a cornerstone of power generation methodologies.

Furthermore, recent statistics released by the National Energy Administration of China [2] underscore China's pivotal role in this global transition. In 2023, China's installed solar power capacity surged to 610 million kW, reflecting an impressive year-on-year growth of 55.2%. This exponential expansion positions China as a frontrunner in the global energy transition efforts. Consequently, photovoltaic power generation stands out as a key trajectory for future development, destined to assume an increasingly significant role in global energy supply dynamics. In comparison to traditional thermal power generation methods, photovoltaic power generation offers distinct advantages such as abundant resource availability and environmentally friendly operation. However, it is highly susceptible to weather fluctuations, resulting in pronounced intermittency and unstable power output. Integrating photovoltaic systems into the power grid often requires prioritizing grid stability over economic efficiency [3], thus presenting a trade-off between system reliability and cost-effectiveness.

As the global energy landscape transitions towards renewable sources, the scale of photovoltaic power generation is poised to significantly expand in the coming years. This inevitable growth poses a substantial challenge to the stability of global power systems. Consequently, finding effective strategies to mitigate or preempt the impact of photovoltaic power generation's intermittency on grid stability has become an urgent necessity. Currently, preempting the adverse impacts of grid-connected photovoltaic power generation often entails forecasting fluctuations in power output beforehand. This proactive approach facilitates efficient capacity allocation, optimal resource utilization, and fortification of power grid safety and stability [4]. Traditional methods for predicting photovoltaic power generation output primarily encompass physical approaches [5] and statistical methodologies [6]. The physical prediction model hinges on understanding the interplay between solar irradiance's physical state and dynamic movement within the atmosphere. It comprises mathematical equations delineating atmospheric conditions. However, this method's accuracy greatly relies on weather forecasts, whereas rapid weather changes can compromise its predictive precision [5]. Conversely, statistical methods leverage historical power generation data and derived information to forecast time series, albeit with limited accuracy [7]. In recent years, propelled by advancements in artificial intelligence, machine learning has emerged as a promising avenue for photovoltaic power generation prediction [6]. Machine-learning models excel in establishing intricate relationships between inputs and outputs, even when these relationships cannot be explicitly defined. This characteristic renders machine learning particularly adept at tasks such as pattern recognition, classification, data mining, and prediction [8].

Given the nonlinear nature of meteorological data and the intricate interplay between various factors influencing photovoltaic power generation, forecasting output represents a quintessential nonlinear problem. Artificial neural networks stand out as a popular choice among machine-learning algorithms due to their ability to model and analyze complex nonlinear processes without necessitating explicit assumptions about the relationship between input and output variables. Consequently, artificial neural networks find widespread application in addressing this challenge. For instance, Regidor et al. [9] investigated the predictive capabilities of an artificial neural network algorithm for short-term photovoltaic power generation output forecasting within the 1–6 h range. This study incorporated temperature, pressure, wind, humidity, and solar irradiance as input variables, achieving outstanding predictive accuracy with a normalized root mean square error (NRMSE) of less than 20%. Similarly, Kumar et al. [10] combined historical power generation data with artificial neural networks to conduct short-term predictions of hourly, daily, and weekly average values of solar photovoltaic power generation. The findings showcased a high

correlation between predicted and actual outputs, underscoring the efficacy of artificial neural networks in addressing various photovoltaic power generation prediction challenges. Oscar et al. [11] developed a measurement Internet of Things prototype to collect on-site voltage and current measurements from the panel, as well as the environmental factors of lighting, temperature, and humidity in the system's proximity. Based on an artificial neural network, the most effective model utilized lighting, temperature, and humidity with an *RMSE* of 0.2553. Collectively, these studies exemplify the extensive utilization of artificial neural networks in the area of photovoltaic power generation prediction, highlighting their ability to deliver optimal prediction outcomes for typical issues encountered in this domain.

Certainly, the impact of winter weather conditions, particularly rain and snow, significantly affects photovoltaic power generation, primarily due to power losses stemming from snow and ice accumulation on photovoltaic panel equipment [12]. Consequently, studying the prediction of photovoltaic power generation under these unique weather conditions is imperative. Several researchers have delved into this area. For instance, Vaz et al. [13] employed artificial neural networks to predict the winter output of photovoltaic systems across five distinct locations in Utrecht, the Netherlands. Their study entailed a month-long prediction test utilizing local meteorological data and measurement data from nearby photovoltaic systems, yielding *NRMSEs* ranging between 9% and 25%. Similarly, Meng et al. [14] focused on predicting the average daily photovoltaic power generation in North China during winter, categorizing winter days into three types: sunny, cloudy, and rainy or snowy. Their findings showcased low prediction errors for sunny and cloudy days, with mean absolute percentage error (*MAPE*) as low as 2.83% and 3.89%, respectively. However, in rainy and snowy weather, *MAPE* increased by 14.29%. Despite these efforts, current research on winter photovoltaic power generation forecasting still faces challenges regarding prediction accuracy, underscoring the need for further exploration and refinement in this domain.

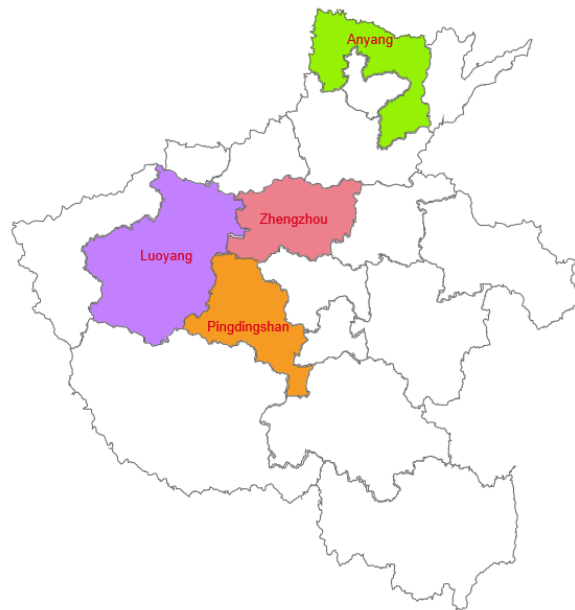
In this study, a short-term photovoltaic power generation prediction model is proposed based on an enhanced BP neural network to analyze photovoltaic power generation output in Henan Province, China. The model is graded into two sub-networks based on weather conditions: no-rain and no-snow, and rainy and snowy. Historical power generation data along with key meteorological variables serve as input parameters for the model. Moreover, the model incorporates rainfall and snowfall amounts as additional input variables. Through this approach, hourly predictions of photovoltaic power generation output in Henan Province for the following day were successfully achieved. The results confirmed the high prediction accuracy of the model. Given the urgent demand for new energy development, this research is of utmost practical significance as it promotes the advancement of photovoltaic power generation technology and facilitates adjustments in the energy structure.

## 2. Methods

### 2.1. Photovoltaic Power Generation Prediction Database Establishment

Henan Province, spanning between  $31^{\circ}23'$  and  $36^{\circ}22'$  north latitude and  $110^{\circ}21'$  and  $116^{\circ}39'$  east longitude, boasts a vast geographical expanse and encompasses 18 prefecture-level cities. Henan Province is considered China's largest province in terms of installed renewable energy capacity and should be an important research target. Given the extensive territory of Henan Province, this study only focused on selected representative cities. These cities include the provincial capital of Zhengzhou ( $34^{\circ}16'$ – $34^{\circ}58'$  north latitude and  $112^{\circ}42'$ – $114^{\circ}14'$  east longitude), the industrial hub of Luoyang ( $34^{\circ}32'$ – $34^{\circ}45'$  north latitude and  $112^{\circ}16'$ – $112^{\circ}37'$  east longitude), and the substantial installed capacity of renewable energy power generation regions of Anyang ( $35^{\circ}12'$ – $36^{\circ}22'$  north latitude and  $113^{\circ}37'$ – $114^{\circ}58'$  east longitude) and Pingdingshan ( $33^{\circ}08'$ – $34^{\circ}20'$  north latitude and  $120^{\circ}14'$ – $133^{\circ}45'$  east longitude). Choosing regions with different latitudes and longitudes to include as many solar radiation conditions as possible can help improve the applicability of the model and reduce prediction errors. The geographic distribution of these selected

cities is depicted in Figure 1. These cities collectively involve diverse weather patterns and varying conditions of photovoltaic power generation installation.



**Figure 1.** The geographical location distribution of representative cities.

Initially, the hourly photovoltaic power generation data for the aforementioned cities spanning from January 2022 to April 2023 was meticulously collected and organized. Leveraging the central latitude and longitude positions of these cities, comprehensive datasets encompassing key meteorological factors, including pressure, temperature, humidity, wind speed, net solar irradiance, and more, were systematically queried. Notably, the meteorological data were available hourly. The photovoltaic power generation performance data and meteorological datasets were then carefully matched and correlated to construct a comprehensive database tailored for forecasting photovoltaic power generation.

## 2.2. Data Correlation Analysis

The output of photovoltaic power generation is not solely dependent on solar irradiance but is also influenced by other meteorological factors such as temperature and humidity. Considering all meteorological factors as inputs to the forecast could overcomplicate the model, resulting in increased computational complexity, which may impede timely predictions. Moreover, it could potentially elevate prediction errors and compromise accuracy. Hence, it is essential to properly choose meteorological factors [15]. In practical photovoltaic power stations, each meteorological factor exerts an influence, albeit with varying degrees of impact. Correlation analysis can examine the relationships between different variables. By conducting correlation analysis between photovoltaic power generation output and various meteorological factors, the magnitude of each factor's impact on power output could be discerned. In this study, the Pearson correlation coefficient was employed to evaluate the correlation between meteorological factors and photovoltaic power generation output, with the calculation formula depicted in Equation (1).

$$r = \frac{cov(X, Y)}{\sigma_X \sigma_Y} \quad (1)$$

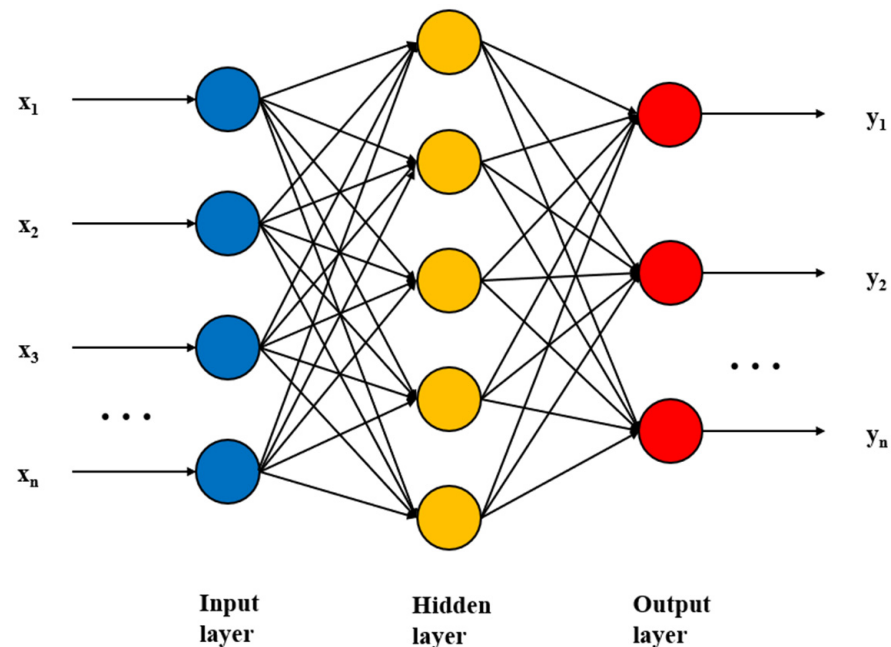
In the above equation,  $cov(X, Y)$  represents the covariance of variables  $X$  and  $Y$ ,  $\sigma_X$  and  $\sigma_Y$  represent the standard deviation of variables  $X$  and  $Y$ , respectively, and  $r$  is the correlation coefficient. A correlation coefficient of 0 indicates no correlation between the two variables. Positive or negative values of  $r$  denote positive or negative correlations, respectively. Generally, when  $|r| < 0.2$ , the variables are considered essentially uncor-

related. When  $0.2 \leq |r| < 0.4$ , the correlation is deemed weak. For  $|r|$  values falling between 0.4 and 0.6, the correlation is moderate, while values between 0.6 and 0.8 indicate a strong correlation. If  $|r| \geq 0.8$ , the variables are highly correlated. Hence, when  $|r|$  between a meteorological factor and photovoltaic power generation output exceeds 0.2, it is considered as a relevant input variable for the model.

### 2.3. GA–BP Neural Network Algorithm

The backpropagation (BP) neural network, proposed in 1986, stands as one of the most prevalent artificial neural networks [16]. It comprises multiple layers and employs the error backpropagation algorithm for training. Generally, the BP neural network leverages the gradient descent method, which employs gradient search techniques to minimize the disparity between actual and expected output values. The BP neural network operation entails two key processes: signal forward propagation and error backpropagation. During forward propagation, input signals traverse through the network's layers, undergoing transformations via hidden layers to produce output signals. When the actual output deviates from the expected output, the error is propagated backward through the network. Error backpropagation involves transferring output errors backward from the output layer to the input layer, distributing errors across all units within each layer. Subsequently, network weights and thresholds are adjusted based on the error signals from each layer, minimizing errors along the gradient direction. Through iterative learning and training, the actual output progressively converges towards the expected output, culminating in the determination of network parameters corresponding to minimal error, thus concluding the training process.

Theoretically, a three-layer neural network is capable of approximating any nonlinear function [17]. Accordingly, this study adopts a BP neural network configuration comprising one input layer, one hidden layer, and one output layer, as depicted in Figure 2.



**Figure 2.** Topology of a BP neural network.

The genetic algorithm (GA) [18] serves as a stochastic global search optimization algorithm, mimicking the evolutionary principle of “survival of the fittest” observed in nature. The fundamental concept revolves around generating a population of individuals from an initial state, with subsequent iterations refining the population through random selection, crossover, and mutation operations. This iterative process drives the population towards regions of higher fitness within the search space. Through successive generations

of reproduction and adaptation, the population gradually converges towards individuals that best suit the environment, ultimately yielding an optimal solution to the problem. However, gradient descent is susceptible to local optima and may struggle to achieve global optimization. By integrating the BP neural network with GA (GA-BP), the shortcomings of falling into local optima and suboptimal prediction model performance can be effectively mitigated. This hybrid approach offers enhanced robustness against local optimality and enhances prediction accuracy.

#### *2.4. The Prediction Model Based on GA-BP Neural Network*

The objective of this study is to predict the photovoltaic power generation output in Henan Province, China, specifically during winter. The unique weather conditions characterized by rain and snow pose challenges, as snow accumulation on photovoltaic panels significantly impacts output, resulting in intricate fluctuations that complicate prediction processes. To streamline the model and enhance accuracy, the forecasting framework was split into two sub-models based on rainy or snowy weather conditions.

In non-rainy and snowy weather scenarios, primary meteorological factors were selected as model inputs through correlation analysis. Conversely, in the sub-model for predicting rain and snow weather; rainfall and snowfall data were integrated to evaluate their potential effects on photovoltaic power generation system performance. Furthermore, photovoltaic power generation is influenced by factors such as the direction and angle of photovoltaic array installation, as well as system conversion efficiency. Neglecting these factors could lead to inaccurate predictions. Given that meteorological factors alone may not adequately capture information from the power generation equipment perspective, historical power generation data were incorporated as inputs. Importantly, since all historical power generation data originated from the same power generation system, they inherently contain key information regarding the system, including installation position and angle [10]. This holistic approach ensured a comprehensive consideration of both meteorological and equipment-related factors, thereby improving prediction accuracy.

The forecasting process unfolded as follows: initially, through correlation analysis, meteorological factors predominantly affecting the photovoltaic power generation output of each city were identified. For the non-rainy and snowy weather forecasting sub-model, input variables encompassed primary meteorological factors, photovoltaic power generation output from the previous day, and key meteorological factors for the forecast day. In the sub-model dedicated to rain and snow weather, input variables additionally included rainfall and snowfall data. The output signified the forecasted photovoltaic power generation output for the forecast day. In predicting rain and snow weather conditions, two comparison routes were adopted: (1) Model I, where the input variables excluded rainfall and snowfall; (2) Model II, where rainfall and snowfall were considered as input variables. By comparing the prediction performance of these two models, the influence of rain and snowfall on the prediction of photovoltaic power generation output could be better recognized.

The photovoltaic power generation output forecasting process is elucidated in Figure 3. The detailed parameter settings for the BP neural network were as follows: the number of hidden layer nodes varied from 1 to 50, and the optimal number of hidden layer nodes was determined through iterative training. For the GA, the initial population size was set to 40, the maximum number of iterations was limited to 100, the individual length was fixed at 20, and the genetic, crossover, and mutation probabilities were set to 0.95, 0.8, and 0.01, respectively.

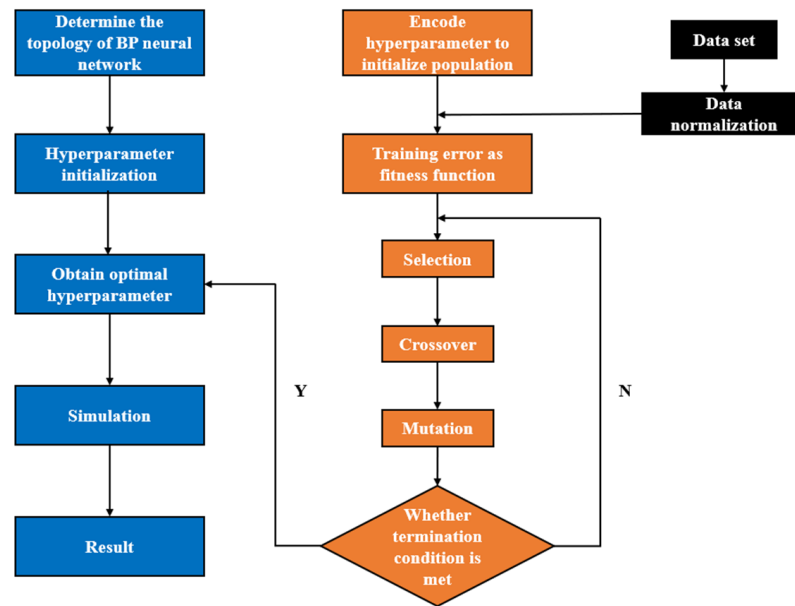


Figure 3. Photovoltaic power output forecasting process.

### 3. Results and Discussion

In this study, data from December 2022 to February 2023 were utilized as training and validation samples for the sub-model aimed at predicting non-rainy and snowy weather. The large sample size ensured the effectiveness of the model. Conversely, for the sub-model focusing on predicting in rain and snow weather conditions, data encompassing January and February 2022 were incorporated to ensure adequate sample quantity. Given the facts that photovoltaic power generation predominantly occurs during daytime when solar irradiance is sufficient and that output diminishes during other times, this study forecast photovoltaic power generation output specifically from 8:00 a.m. to 4:00 p.m. daily.

#### 3.1. Photovoltaic Power Generation Output Forecasting of Non-Rainy and Snowy Weather

In order to select major meteorological factors as inputs, a correlation analysis was conducted between photovoltaic power generation output and various meteorological factors in the database. The results of this analysis are presented in Table 1.

Table 1. Correlation analysis results of main meteorological factors.

City	Anyang	Pingdingshan	Luoyang	Zhengzhou
Surface pressure (hPa)	0.11	0.037	−0.015	0.077
Temperature (°C)	0.43	0.52	0.43	0.51
Relative humidity (%)	−0.49	−0.62	−0.49	−0.49
Warp wind speed (m/s)	0.23	0.22	−0.065	0.18
Zonal wind speed (m/s)	−0.03	0.31	0.17	0.13
Net solar irradiance (J/m <sup>2</sup> )	0.75	0.77	0.75	0.83

The data presented in Table 1 suggest that the photovoltaic power generation performance of the four cities was primarily influenced by temperature, humidity, and net solar irradiance. It is worth noting that the most critical factor was net solar irradiance, with a correlation coefficient of over 0.75. Additionally, temperature and humidity were relatively major influencing factors, exhibiting correlation coefficients ranging from 0.4 to 0.6. Conversely, the effects of other meteorological factors may be deemed relatively negligible. Therefore, in the sub-model dedicated to non-rainy and snowy weather, the principal identified meteorological factors were net solar irradiance, temperature, and humidity.

The model’s verification utilized data spanning from 26 February 2023 to 28 February 2023. The forecasted photovoltaic power generation output for the four cities during winter’s

non-rainy and snowy weather is depicted in Figures 4–7. It was observed that the predicted output closely followed the trend of the actual output, with overall differences being insignificant. However, slight deviations were noticeable near the maximum daily photovoltaic power generation output, a phenomenon that is also documented in the literature [19]. This discrepancy could be attributed to the relatively sparse data points near the maximum value, which somewhat diminishes prediction accuracy. To comprehensively assess the prediction performance of the model, error indices such as root mean square error (*RMSE*), *MAPE*, *NRMSE*, and  $R^2$  were employed. The calculation equations for these indices are provided in Equations (2)–(5).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2} \quad (2)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - y'_i}{y_i} \right| \quad (3)$$

$$NRMSE = RMSE / (y_{max} - y_{min}) \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y'_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

As illustrated in Table 2, the *RMSE* for Anyang and Pingdingshan City ranged between 80 and 90 MW, while for Luoyang City, it fell between 60 and 70 MW. Additionally, the *RMSE* for Zhengzhou City was below 40 MW. This discrepancy could be attributed to the fact that the maximal photovoltaic power generation output in Anyang and Pingdingshan City was double that of Zhengzhou City, leading to a higher *RMSE*. Furthermore, *MAPE* ranged between 9.66% and 12.61%, indicating that the predicted value had an error of no more than 13% relative to the actual value. Comparing to that in recent similar research [19], the obtained *MAPE* value seems 7% bigger. However, this study was aimed at predicting the power generation output of all photovoltaic power stations in a city (hundreds or even thousands of megawatts) rather than just a specific power station (tens of megawatts). Thus, this discrepancy may be attributed to the presence of various peak values and non-stationary components in historical power generation data of dozens of photovoltaic power stations, coupled with uncertainty and variability of the corresponding meteorological conditions that significantly impact power generation.

**Table 2.** Photovoltaic power generation output forecasting error of non-rainy and snowy weather.

City	<i>RMSE</i>	<i>MAPE</i>	<i>NRMSE</i>	$R^2$
Anyang	89.41	10.67%	7.45%	0.9359
Pingdingshan	84.47	12.61%	6.97%	0.9430
Luoyang	64.91	11.98%	6.69%	0.9487
Zhengzhou	37.76	9.66%	8.91%	0.9220

Generally, *NRMSE* can compare prediction errors from different samples. *NRMSE* in this work fell within the range of 7–9%, demonstrating a positive outcome. Compared with those in the literature [20], despite not further subdividing no-rain and snow weather, relatively accurate predictions were achieved in this study. *NRMSE* decreased by about 8%. Obviously, the selection of major meteorological elements through correlation analysis and the consideration of historical generation data helped improve the model's forecasting performance. Moreover,  $R^2$  exceeded 92%, indicating good generalization performance of the model.

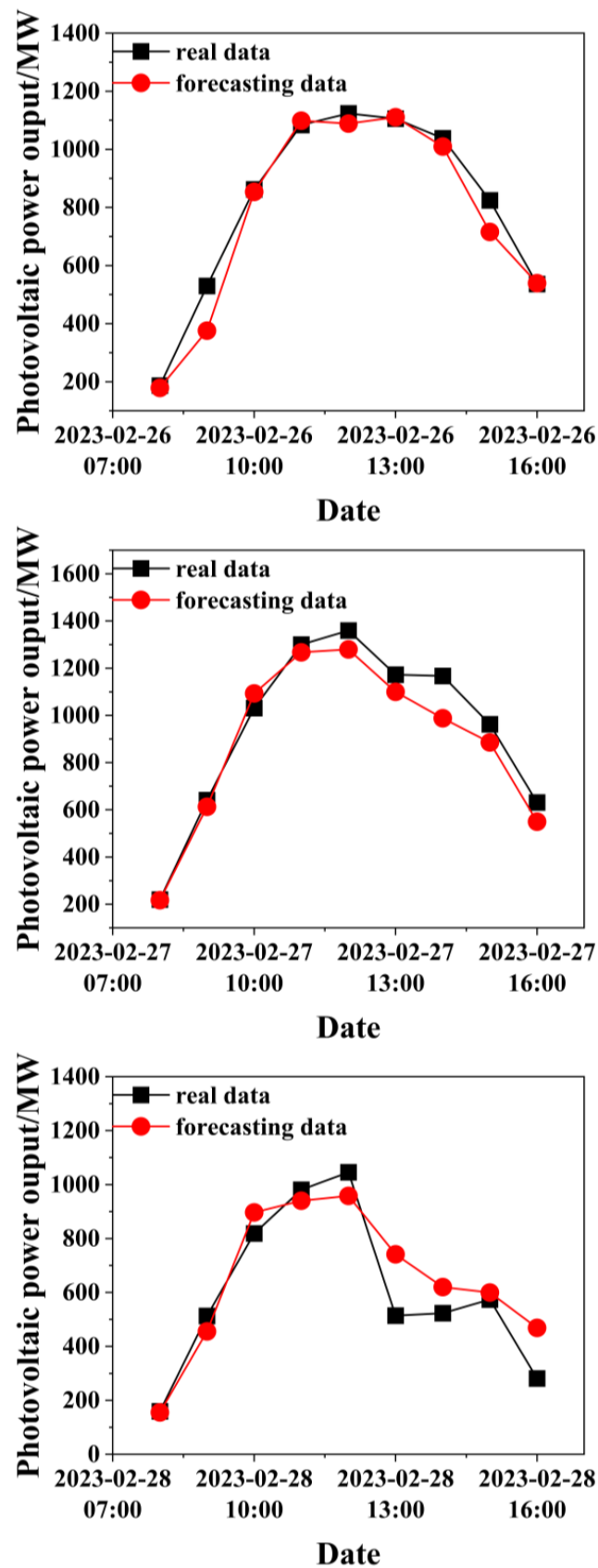


Figure 4. The forecasting results in non-rain and snow weather of Anyang city (photovoltaic power output versus time of February 26, 27, and 28, 2023 from 8 a.m. to 4 p.m.).

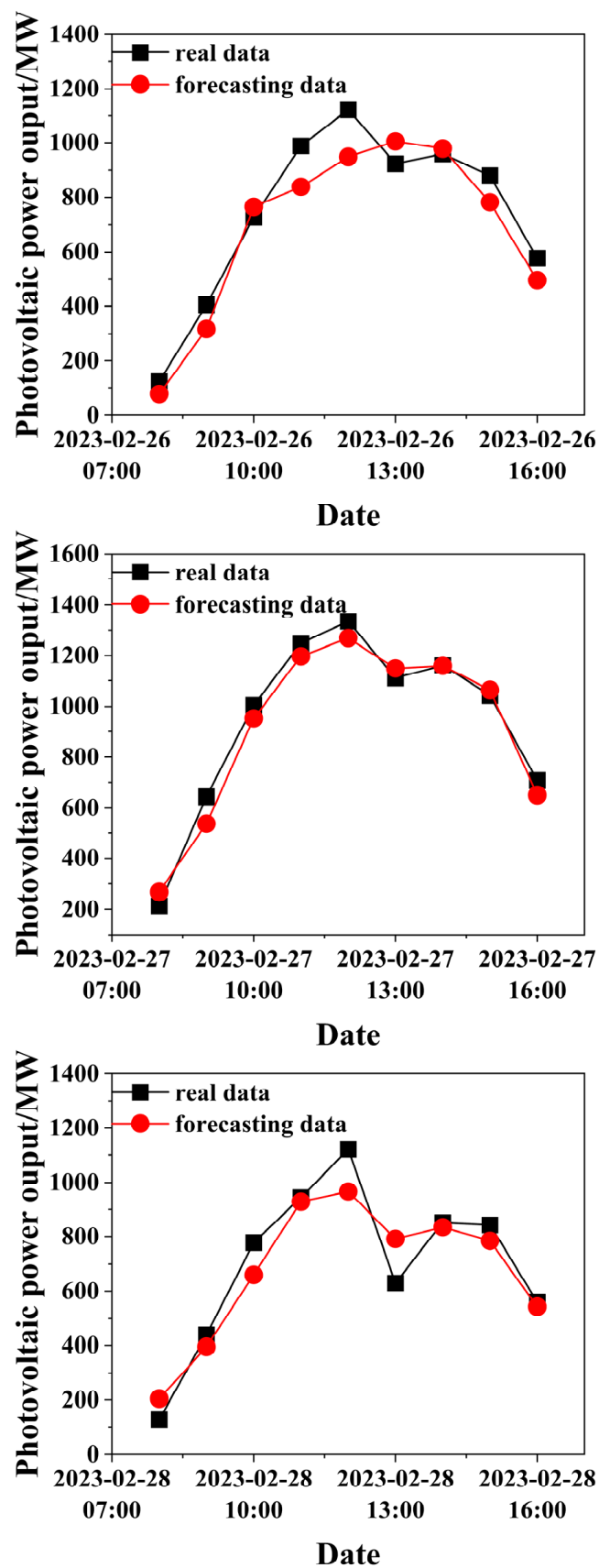


Figure 5. The forecasting results in non-rain and snow weather of Pingdingshan city (photovoltaic power output versus time of February 26, 27, and 28, 2023 from 8 a.m. to 4 p.m.).

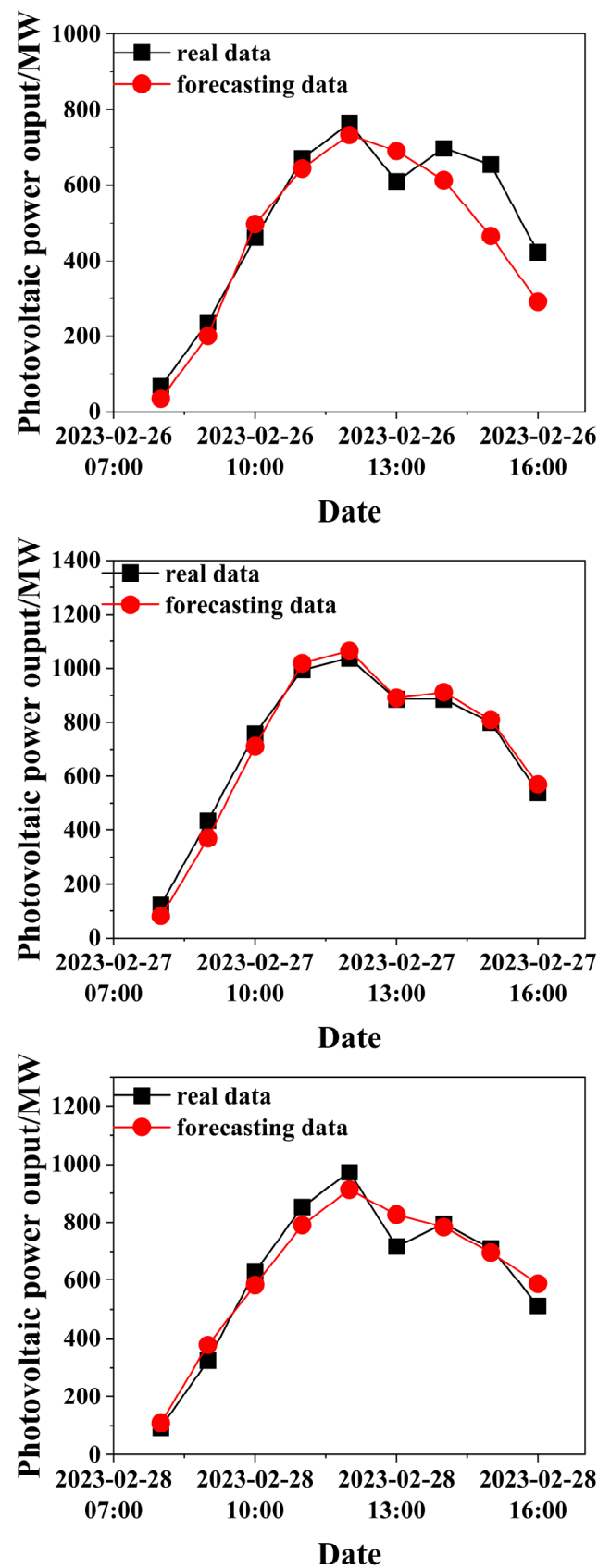


Figure 6. The forecasting results in non-rain and snow weather of Luoyang city (photovoltaic power output versus time of February 26, 27, and 28, 2023 from 8 a.m. to 4 p.m.).

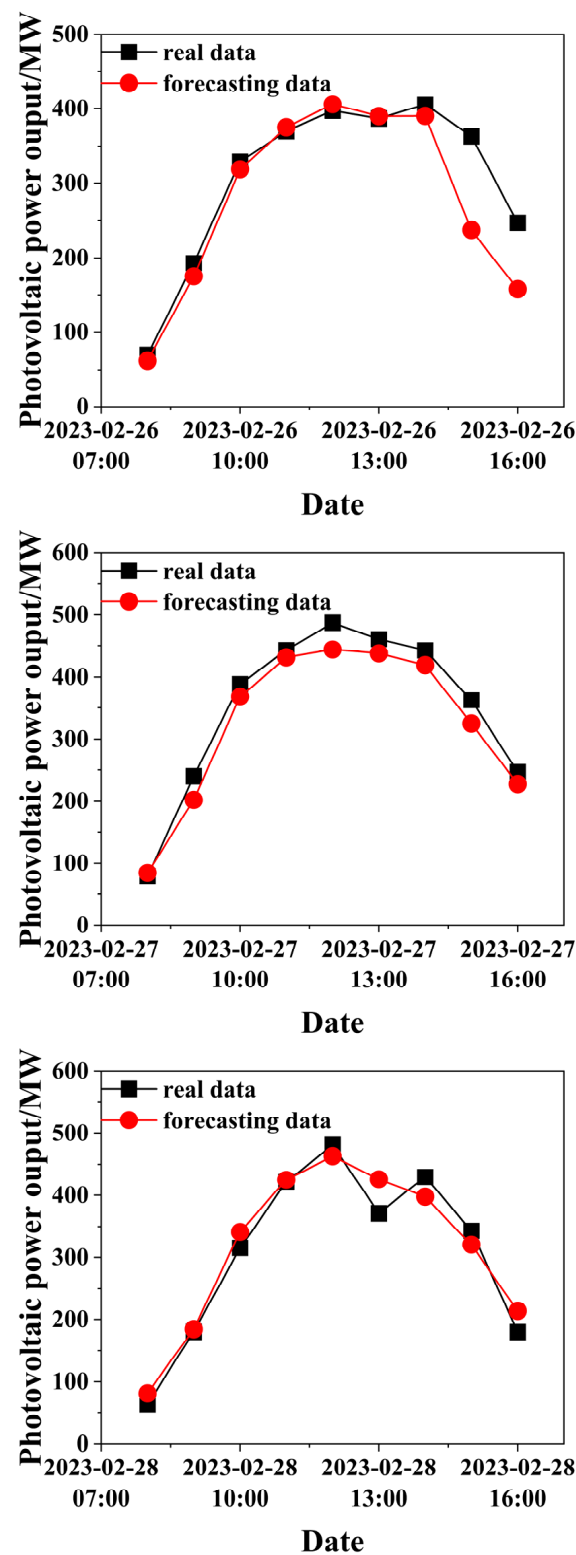


Figure 7. The forecasting results in non-rain and snow weather of Zhengzhou city (photovoltaic power output versus time of February 26, 27, and 28, 2023 from 8 a.m. to 4 p.m.).

In Equations (2)–(5),  $y_i$  represents the actual value,  $y'_i$  represents the predicted value, and  $y_{max}$  and  $y_{min}$  represent the maximum and minimum values of the sample, respectively. *RMSE* assesses the average deviation between predicted and actual values, while *MAPE* measures the percentage error between the predicted and actual values. *NRMSE* enables the

comparison of error magnitudes across different samples, and  $R^2$  describes the goodness of fit.

In summary, the sub-model for forecasting non-rainy and snowy weather achieved satisfactory results. The prediction error of different cities was at similar levels, and the small differences could be attributed to different geographic locations and power generation site information. Consequently, the proposed model is applicable to photovoltaic power generation systems in diverse situations.

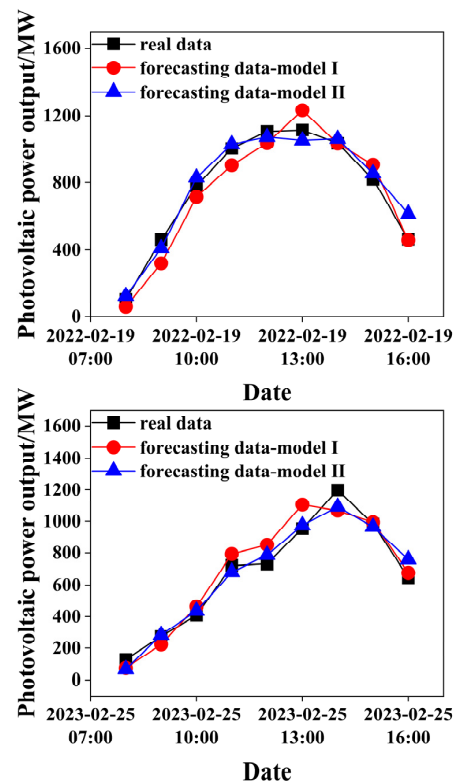
### 3.2. Photovoltaic Power Generation Output Forecasting of Rain and Snow Weather

Due to the influence of rain and snow, photovoltaic panels may be obstructed by snow and ice, thereby impeding photovoltaic power generation output [12]. Consequently, it becomes imperative to conduct photovoltaic power generation output prediction during rainy and snowy conditions. Table 3 illustrates the correlation analysis results of photovoltaic power generation output with rainfall and snowfall in each city. It is evident that rain and snow do exert a discernible impact on photovoltaic power generation output, with varying degrees of influence observed across different cities.

**Table 3.** Correlation analysis results of rain and snow.

City	Rainfall	Snowfall
Anyang	−0.27	−0.25
Pingdingshan	−0.23	−0.24
Luoyang	−0.14	−0.13
Zhengzhou	−0.14	−0.12

The forecasting results of photovoltaic power generation output in four cities during winter rain and snow conditions are depicted in Figures 8–11. The model was validated using data from the last day of continuous rain and snow weather in 2022 and 2023, respectively.



**Figure 8.** The forecasting results in rain and snow weather of Anyang city (photovoltaic power output versus time of February 19, 2022 and February 25, 2023 from 8 a.m. to 4 p.m.).

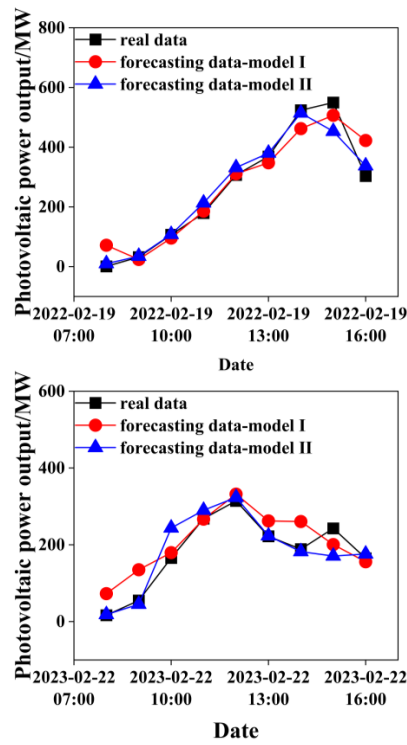


Figure 9. The forecasting results in rain and snow weather of Pingdingshan city (photovoltaic power output versus time of February 19, 2022 and February 22, 2023 from 8 a.m. to 4 p.m.).

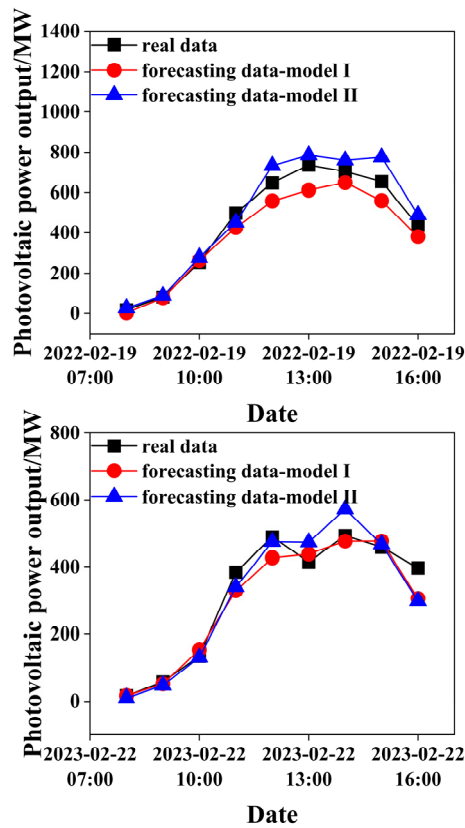
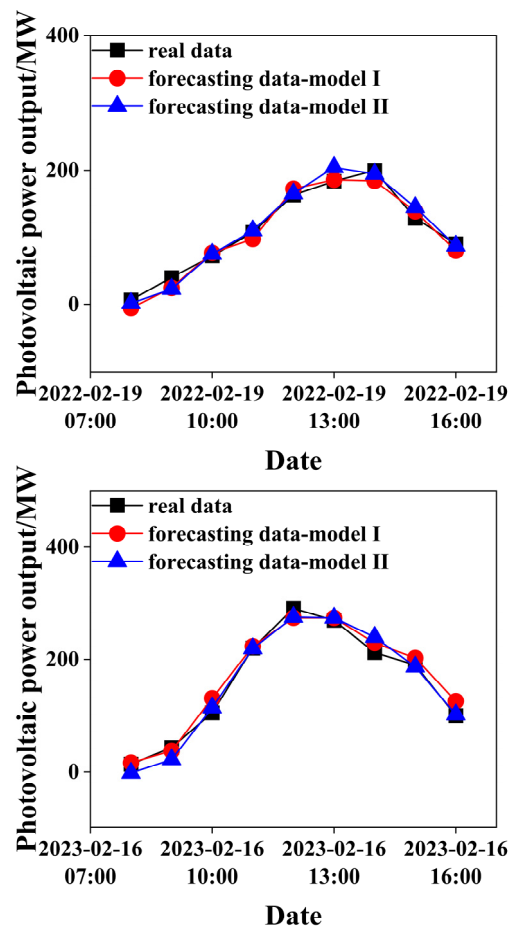


Figure 10. The forecasting results in rain and snow weather of Luoyang city (photovoltaic power output versus time of February 19, 2022 and February 22, 2023 from 8 a.m. to 4 p.m.).



**Figure 11.** The forecasting results in rain and snow weather of Zhengzhou city (photovoltaic power output versus time of February 19, 2022 and February 16, 2023 from 8 a.m. to 4 p.m.).

It was observed that the predicted output of both Model I and Model II generally aligned with the actual output. However, noticeable discrepancies between predicted and actual values were evident for Pingdingshan and Luoyang cities during certain periods. This disparity may have been due to two main factors: Firstly, the meteorological conditions represented by the longitude and latitude coordinates of the city center may not fully reflect the actual conditions [21]. Accurate meteorological data are pivotal for precise weather predictions. Secondly, rainfall and snowfall can only partially depict the extent to which photovoltaic power generation output was influenced by snow and ice. The specific impact should be determined based on on-site information. These disparities may impede the full exploitation of the actual change pattern in learning and training of neural networks, thereby affecting prediction accuracy.

Furthermore, Model II exhibited lower prediction deviation compared to that of Model I, and its prediction curve more closely resembled the actual output curve. This suggests that the inclusion of rainfall and snowfall in Model II led to enhanced prediction performance. Obviously, it will be more complicated when considering the prediction of photovoltaic power generation under extreme weather of rain and snow.

Table 4 displays the forecast error of photovoltaic power generation output during rain and snow weather. Compared with that of the non-rainy and snowy weather sub-model, the *RMSE* was lower because photovoltaic power generation was hindered in rainy and snowy conditions, leading to a significant decline in power generation, which in turn reduced the forecast deviation. In general, the larger the size of the predicted object, the larger the *RMSE* [11,20]. Similarly, *NRMSE* also fell down. However, *MAPE* seemed to have grown because in extreme weather there were times when the output of power generation

was very small, which magnified the error when measured by percentage error. For the high *MAPE* in Pingdingshan city, it was found that the prediction deviation in the city was quite significant at low output levels below 20MW, resulting in an amplification of *MAPE*.

**Table 4.** Photovoltaic power generation output forecasting error in rain and snow weather.

City	Model	<i>RMSE</i>	<i>MAPE</i>	<i>NRMSE</i>	<i>R</i> <sup>2</sup>
Anyang	Model I	86.04	14.19%	7.85%	0.9339
	Model II	61.83	10.52%	5.64%	0.9965
Pingdingshan	Model I	49.54	469.78%	9.03%	0.8955
	Model II	37.02	67.54%	6.75%	0.9426
Luoyang	Model I	58.21	15.08%	8.08%	0.9365
	Model II	55.35	14.89%	7.68%	0.9668
Zhengzhou	Model I	13.26	20.91%	4.67%	0.9743
	Model II	12.58	19.41%	4.43%	0.9818

Additionally, improving the model also improved its performance, which suggested that adding rainfall and snowfall was essential for accurate prediction. *RMSE* was reduced by an average of 13 MW for Anyang and Pingdingshan city, but only by about 2 MW for Luoyang and Zhengzhou city. This can be explained by the greater negative correlation between power generation output and rainfall and snowfall in Anyang and Pingdingshan city. For *MAPE*, all cities except Pingdingshan had errors of less than 20%, which aligned with the findings of Meng et al. [14] and satisfied the power grid's requirements. This was because a similar input/output pattern was used. It was worth noting that this study aimed to predict hourly output, while Meng et al. [14] aimed to predict daily power generation. The prediction level improved, but the prediction error was still relatively small. Compared with that in the literature [21], *MAPE* had an increase of 3–5% due to the consideration of rainfall and snowfall as model inputs.

In addition, the *NRMSE* for each city was below 9%, which was a significant improvement compared to the results of Vaz et al. [13], signifying the accuracy of the model. In their study, the rainfall and snowfall were not taken into account separately. Compared with that found by Meng et al., *NRMSE* still had a reduction of 2–4%. This further shows that it is not enough to forecast only on the basis of weather type classification, and the impact of rain and snow on power generation equipment also needs to be included. Consequently, *R*<sup>2</sup> exceeded 0.9, also validating the model performance.

To sum up, the sub-model for forecasting rainy and snowy weather achieved satisfactory results. By adding rainfall and snowfall as model inputs, the model was optimized and its performance increased. Meanwhile, the effect of the model on four cities in different geographical locations was at a similar level. However, it is obvious that the greater the correlation between rainfall and snowfall and output, the greater the optimization of the results obtained after adding input. Consequently, the proposed model is applicable to photovoltaic power generation systems in special weather like rain and snow.

#### 4. Conclusions

For photovoltaic power generation, weather disturbances pose a significant challenge when forecasting the output, while forecasting performance holds significant importance for power grid operation and dispatching. In this study, a short-term forecast model for China's Henan Province's winter photovoltaic power generation output was developed based on a genetic algorithm–backpropagation neural network. Firstly, through correlation analysis, the main meteorological factors affecting the power generation output were selected. Then, the historical power generation data were also adopted as inputs to account for device information, and hourly photovoltaic power generation output served as the output. In order to reduce model complexity and improve operation efficiency, the model

was partitioned into two sub-models tailored to different weather types: non-rainy and snowy, and rain and snow. At last, conducting predictions for four representative cities of different latitudes and longitudes in Henan Province, China, the following conclusions could be drawn:

- (1) Through correlation analysis, the primary meteorological factors influencing photovoltaic power generation output were successfully identified. They were solar irradiance (correlation coefficient  $> 0.75$ ), temperature, and humidity ( $0.4 < \text{correlation coefficient} < 0.6$ ). Rainfall and snowfall also played a role (correlation coefficient up to 0.27). By integrating these factors with historical power generation data, a forecasting model for photovoltaic power generation output in winter was developed.
- (2) In the non-rainy and snowy sub-model, *RMSE* ranged between 37.76 and 89.41 MW depending on the magnitude of the predicted object. *MAPE* remained below 13%, meeting the requirement of  $< 20\%$ . Additionally, *NRMSE* fell within the range of 7–9%, and  $R^2$  exceeded 0.92. The errors were in a good range. By selecting major meteorological elements through correlation analysis and the consideration of historical generation data, the model could be successfully applied to power generation prediction in different situations (geographical location, power generation equipment information, weather information).
- (3) During rain and snow weather conditions, model performance was improved by taking into account the effects of rainfall and snowfall. *RMSE* ranged between 12.58 and 61.83. *MAPE* increased slightly but still remained below 20%. Moreover, *NRMSE* was less than 8%, and  $R^2$  exceeded 0.94. These could demonstrate the model's effectiveness in predicting photovoltaic power output even under challenging weather conditions. In addition, the difference in the forecasting of these four cities also indicated that different regions were affected by rain and snow to different degrees, and the greater the impact, the more necessary it was to further consider it.

In future work, more factors will be taken into account to make the prediction model more comprehensive. Specifically, Henan province in the north of China is severely affected by smog. The deposition of flying dust on photovoltaic panels could reduce the output performance of photovoltaic power generation [22,23]. Therefore, the measurement of flying dust and consideration of the effects of different categories will make the model more practical.

**Author Contributions:** Writing—original draft, D.X.; Writing—review & editing, L.L., B.Z., M.L., C.W., Z.G., A.A.S., L.J. and S.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

**Acknowledgments:** The computation was completed in the HPC Platform of Huazhong University of Science and Technology. The authors would also like to thank Guiying Lin for her valuable advice.

**Conflicts of Interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

1. IEA. Snapshot of Global Photovoltaic Markets. April 2023. Available online: <https://iea-pvps.org/snapshot-reports/snapshot-2023> (accessed on 30 April 2023).
2. National Energy Administration. Available online: [https://www.nea.gov.cn/2024-01/26/c\\_1310762246.htm](https://www.nea.gov.cn/2024-01/26/c_1310762246.htm) (accessed on 26 January 2024).
3. Chakravarthi, B.N.C.V.; Rao, G.V.S.K. Impact of Power Quality Issues in Grid Connected photovoltaic System. In Proceedings of the 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 5–7 November 2020; pp. 155–158.

4. Das, U.K.; Tey, K.S.; Seyedmahmoudian, M. Forecasting of photovoltaic power generation and model optimization: A review. *Renew. Sustain. Energy Rev.* **2018**, *81*, 912–928. [[CrossRef](#)]
5. Mayer, M.J.; Gróf, G. Extensive comparison of physical models for photovoltaic power forecasting. *Appl. Energy* **2021**, *283*, 116239. [[CrossRef](#)]
6. Akhter, M.N.; Mekhilef, S.; Mokhlis, H. Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques. *IET Renew. Power Gener.* **2019**, *13*, 1009–1023. [[CrossRef](#)]
7. Mutaz, A.; Csaba, C. Evaluating neural network and linear regression photovoltaic power forecasting models based on different input methods. *Energy Rep.* **2021**, *7*, 77601–77614.
8. Voyant, C.; Notton, G.; Kalogirou, S. Machine learning methods for solar irradiance forecasting: A review. *Renew. Energy* **2017**, *105*, 569–582.
9. Regidor, M.; Pilar, M.; Sancho, M. Forecasting short-term solar irradiance based on artificial neural networks and data from neighboring meteorological stations. *Sol. Energy* **2016**, *134*, 119–131.
10. Kumar, P.M.; Saravanakumar, R.; Karthick, A. Artificial neural network-based output power prediction of grid-connected semitransparent photovoltaic system. *Environ. Sci. Pollut. Res.* **2022**, *29*, 10173–10182. [[CrossRef](#)] [[PubMed](#)]
11. Lobato-Nostroza, O.; Chavez-Campos, G.M.; Morales-Cervantes, A. Predictive Modeling of Photovoltaic Panel Power Production through On-Site Environmental and Electrical Measurements Using Artificial Neural Network. *Metrology* **2023**, *3*, 347–364. [[CrossRef](#)]
12. Pawluk, R.E.; Chen, Y.; She, Y. photovoltaic electricity generation loss due to snow—A literature review on influence factors, estimation, and mitigation. *Renew. Sustain. Energy Rev.* **2019**, *107*, 171–182. [[CrossRef](#)]
13. Vaz, A.G.R.; Elsinga, B.; Van Sark, W. An artificial neural network to assess the impact of neighbouring photovoltaic systems in power forecasting in Utrecht, the Netherlands. *Renew. Energy* **2016**, *85*, 631–641. [[CrossRef](#)]
14. Meng, M.; Song, C. Daily photovoltaic power generation forecasting model based on random forest algorithm for north China in winter. *Sustainability* **2020**, *12*, 2247. [[CrossRef](#)]
15. Skiba, M.; Mrówczyńska, M.; Bazan-Krzywoszańska, A. Modeling the economic dependence between town development policy and increasing energy effectiveness with neural networks. Case study: The town of Zielona Góra. *Appl. Energy* **2017**, *188*, 356–366. [[CrossRef](#)]
16. Miao, Y.; Lau, S.S.Y.; Lo, K.K.N.; Song, Y.; Lai, H.; Zhang, J.; Tao, Y.; Fan, Y. Harnessing climate variables for predicting PV power output: A backpropagation neural network analysis in a subtropical climate region. *Sol. Energy* **2023**, *264*, 111979. [[CrossRef](#)]
17. Lu, L.; Jin, P.; Pang, G. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nat. Mach. Intell.* **2021**, *3*, 218–229. [[CrossRef](#)]
18. Lambora, A.; Gupta, K.; Chopra, K. Genetic algorithm—A literature review. In Proceedings of the 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 14–16 February 2019; pp. 380–384.
19. Li, Y.; Zhou, L.; Gao, P.; Yang, B.; Han, Y.; Lian, C. Short-term power generation forecasting of a photovoltaic based on PSO-BP and GA-BP neural network. *Front. Energy Res.* **2022**, *9*, 824691. [[CrossRef](#)]
20. Ren, Q.; Zhang, N.; Song, Q.; Li, J. Short-Term PV Power Forecasting Based on BP Network Optimized by Logistic Improved ASO. In Proceedings of the 2021 6th International Conference on Power and Renewable Energy (ICPRE), Shanghai, China, 17–20 September 2021; pp. 1135–1139.
21. Zhang, N.; Wang, S.; Liu, G.C. All-factor short-term photovoltaic output power forecast. *IET Renew. Power Gener.* **2022**, *16*, 148–158. [[CrossRef](#)]
22. Wan, L.; Zhao, L.; Xu, W. Dust deposition on the photovoltaic panel: A comprehensive survey on mechanisms, effects, mathematical modeling, cleaning methods, and monitoring systems. *Sol. Energy* **2024**, *268*, 112300. [[CrossRef](#)]
23. Almukhtar, H.; Lie, T.; Al-Shohani, W. Comprehensive review of dust properties and their influence on photovoltaic systems: Electrical, optical, thermal models and experimentation techniques. *Energies* **2023**, *16*, 3401. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.