

Article

# Hour-Ahead Photovoltaic Power Forecasting Using an Analog Plus Neural Network Ensemble Method

Jingyue Wang <sup>1</sup>, Zheng Qian <sup>1,\*</sup>, Jingyi Wang <sup>1,†</sup> and Yan Pei <sup>2</sup>

<sup>1</sup> School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100191, China; w\_jyue@buaa.edu.cn (J.W.); wangjingyi@cn.fujitsu.com (J.W.)

<sup>2</sup> State Key Laboratory of Operation and Control of Renewable Energy & Storage Systems, China Electric Power Research Institute, Beijing 100192, China; peiyan@epri.sgcc.com.cn

\* Correspondence: qianzheng@buaa.edu.cn

† This Author's Current Affiliation is Fujitsu Research & Development Center Co., Ltd., Beijing 100025, China.

Received: 23 May 2020; Accepted: 22 June 2020; Published: 24 June 2020



**Abstract:** The common analog approach and ensemble methods in photovoltaic (PV) power forecasting are based on the forecasts from several numerical weather prediction (NWP) models. These may be not applicable to the very-short-term PV power forecasting, since forecasts based on NWP models are reliable in horizons longer than six hours. In this paper, a methodology for one-hour-ahead PV power forecasting is proposed. Instead of the NWP models, the persistence method is applied in the analog approach to produce meteorological forecasts. The historical data with meteorological predictions similar to the target forecast hour are identified to train the forecast model. Then, the feed forward neural networks (FNNs) act as the base predictors of the neural network ensemble method to replace the NWP-based PV power prediction methods. The forecast results produced by the FNNs are combined by the random forest (RF) algorithm. The performance of the proposed method is evaluated on a real grid-connected PV plant located in Southeast China. Results show that the proposed method outperforms six benchmark models: the persistence model, the support vector regression (SVR) model, the linear regression model, the RF model, the gradient boosting model, and XGBoost model. The improvements reach up to over 40% for the standard error metrics.

**Keywords:** analog approach; neural network ensemble; photovoltaic power forecasting; persistence method; feed forward neural networks

## 1. Introduction

Fast growing penetration of solar energy has introduced noticeable challenges to the operation of the electric grid due to the dynamic nature and intermittency of solar power. Accordingly, the uncertainty and fluctuations of solar photovoltaic (PV) power must be handled properly. Solar PV output power forecasting has emerged as an efficient tool to address this issue. An accurate power forecasting model can reduce the impacts of PV power variability on the grid, improve system reliability and power quality, and promote large-scale PV power penetration [1,2].

A good number of studies have been conducted to forecast PV power at different temporal scales. Based on the forecast horizon, PV power forecasting methods can be divided into the following categories: very-short-term (i.e., few seconds to hours ahead), short-term (i.e., one to three days ahead), medium-term (i.e., one week ahead), and long-term (i.e., months to year ahead) [3]. In this paper, we focus on the very-short-term PV power forecasting. It is used for forecasting ramps and frequent fluctuations in energy production to ensure unit commitment, scheduling, and dispatching of electrical power [1,4].

Statistical approaches based on historical data have been the most popular technique for PV power forecast [5]. The generally used statistical techniques include the time series methods [6], regression methods [7], and machine learning methods [8]. The considered historical data contain meteorological and power measurements. The main drawback of these methods is the lack of physical information of PV systems during forecasting [5,9,10].

In order to take the physical properties of PV modules into consideration when using statistical techniques for PV power forecast, the analog approach is widely utilized [5,11–16]. As PV output strongly depends on meteorological factors, PV modules would have similar output in similar meteorological conditions [11]. In other words, there will be an approximate relation between the inputs and the forecasted results when the external conditions are similar. Accordingly, the analog methods attempt to select data, which are collected at the conditions similar to that of the target forecast time, to train the forecast models.

To this end, the meteorological predictions on the target forecast time are compared with their historical forecasts. The specific time periods, which have meteorological forecasts analogous to the current meteorological forecasts, are identified. In general, numerical weather prediction (NWP) models are applied to give the predictions [5,11,13–15].

The ensemble method is a technique in machine learning. Through combining several base learners, the ensemble method can achieve higher predictive accuracy than using base learner only. At present, the ensemble method has been widely used in many fields. In recent years, the ensemble method for solar PV power forecasting has become one of the most widely utilized approaches, since it has better predictive performance compared to a single model. In a review paper [9], the ensemble method was identified as an important direction for future solar PV power forecasting research. The commonly used ensemble technique in PV power forecasting is used by blending NWP data from multiple sources [5,11,17–20].

However, these methods may not be applicable to the very-short-term PV power forecasting. NWP simulations over a sizable domain are often delayed, since they run on supercomputers, and require spin-up time. Hence, the NWP-based methods are more accurate with a horizon longer than six hours [5,14,17,21].

In this paper, we propose a method for the very-short-term PV power forecasting. Specifically, the forecast horizon considered is one-hour ahead. This paper makes the following research contributions:

- (1) A novel analog approach for the very-short-term PV power forecasting is proposed to select training data for the target forecast time. The meteorological forecasts and astronomical data are utilized as the features for the analog approach to measure the similarity of the operating conditions. Without using NWP forecasts, the meteorological forecasts are produced by the persistence method, which is simple to use, and has high accuracies for very short timeframe forecasts [5].
- (2) Then an ensemble forecast framework is trained for the target forecast time using the selected training data. It does not utilize NWP models, but rather uses the artificial neural network (ANN) to act as the base predictors, which is more accurate than the NWP-based methods for the forecast horizons shorter than six hours [18]. These approaches have been known as the neural network ensemble (NNE) [22]. In this work, the random forest (RF) algorithm is then adopted to blend the forecasts of the ANNs.

The rest of the paper is organized as follows: Section 2 describes the PV plant and the prepared data for simulating the forecast models. Section 3 discusses the proposed methodology for the very-short-term PV power prediction. In Section 4, numerical results and discussions are presented. Finally, Section 5 concludes the paper.

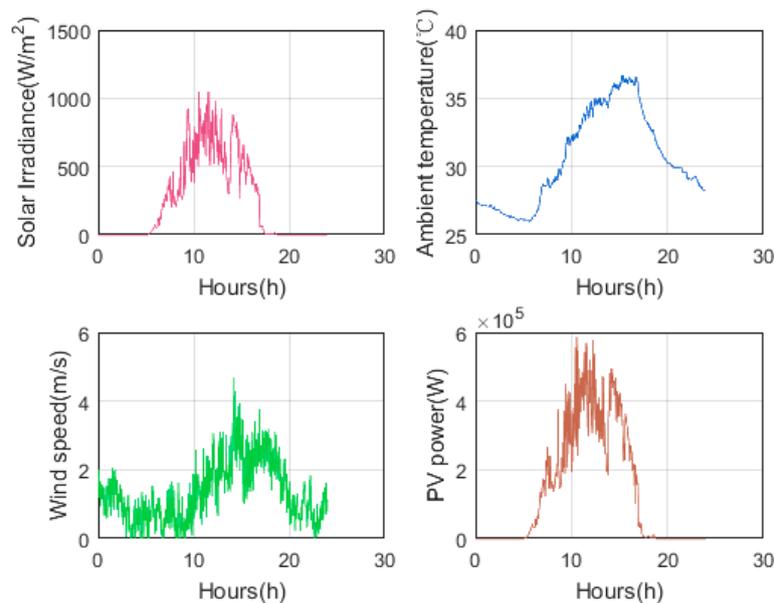
## 2. PV Plant and Data Preparation

In this work, the used data are from a grid-connected PV plant located in Southeast China, which was also analyzed in our previous work [23]. There are 17 sub-arrays in the PV plant, and each sub-array is connected to one inverter. One of the sub-arrays was randomly selected to provide data for the simulation of the method.

The data of the PV plants include the observations of the meteorological variables (i.e., solar irradiance, ambient temperature, and wind speed), and the output power of the PV modules. The collected data range from 24 June 2015 to 23 June 2016. All the measurements are averaged and sampled every minute. Then the one minute measurements are averaged over one hour for one-hour-ahead PV power forecasting. There is no missing point among the one-hour data. The details of raw data without preprocessing are shown in Table 1. Figure 1 shows the randomly selected data which are from 24 June 2015 00:00:00 to 24 June 2015 23:59:00 as the samples.

**Table 1.** Basic statistics of raw data.

| Meteorological Variables                   | Max       | Min | Mean     |
|--|-----------|-----|----------|
| Solar irradiance ( $\text{W}/\text{m}^2$ ) | 1148.0    | 0   | 111.1045 |
| Ambient temperature ( $^{\circ}\text{C}$ ) | 37.3300   | 0   | 20.1984  |
| Wind speed (m/s)                           | 8.2700    | 0   | 1.4251   |
| Photovoltaic (PV) power (W)                | 659,600.0 | 0   | 68,799.0 |



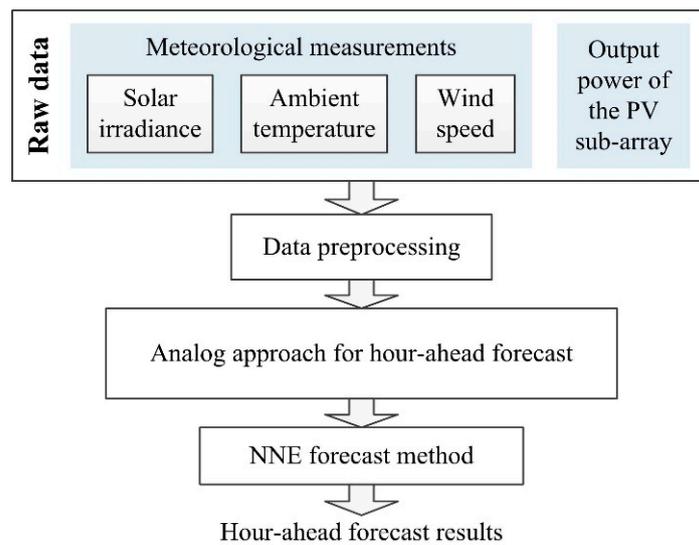
**Figure 1.** Samples of the raw data from 24 June 2015 00:00:00 to 24 June 2015 23:59:00.

## 3. Methods

In this section, the proposed PV power forecasting method is introduced.

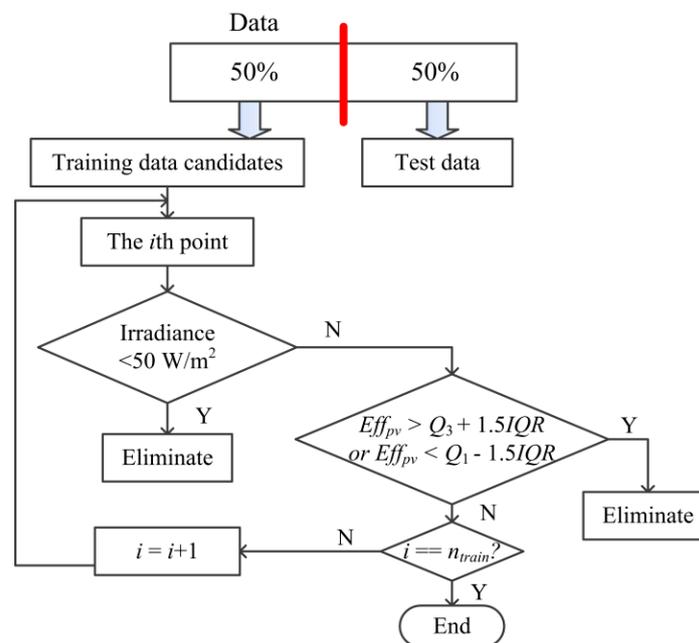
### 3.1. Overview

Figure 2 shows the main procedure of the proposed method, which mainly consists of three steps.



**Figure 2.** Overview of the proposed analog ensemble PV power forecasting method for hour-ahead forecast.

Step (1) Data preprocessing: The flowchart of the data preprocessing is shown in Figure 3. Firstly, the whole data is sequentially split into two subsets. One containing the first 50% data is used as the initial training data candidates. The other one including the remaining 50% data is the test data for validating the proposed method. The specific dates of training data are from 23 June 2015 14:24:00 to 9 January 2016 19:00:00, and the specific dates of test data are from 9 January 2016 20:00:00 to 20 April 2016 23:00:00.



**Figure 3.** Flowchart of the data preprocessing.

Then the observations measured at low solar irradiance levels, which here are arbitrarily set to be less than  $50 \text{ W/m}^2$ , are eliminated from the training data candidates. This is because the uncertainty levels of the observations increase as the solar irradiance levels decrease [24]. In addition, observations with zero power output are also removed.

Under normal conditions, the conversion efficiency of PV modules varies within a range. The PV efficiency may exceed the normal range when the measured PV output power is significantly lower or higher than it should be at the measured irradiance level. What causes the abnormal efficiency may be system failures, inaccurate measurements, data transmission or storage error, etc. The abnormal PV efficiency can be identified using the box plot rule, and then can be removed.

The box plot rule classifies an observation as an outlier if  $Eff_{pv} > Q_3 + 1.5IQR$  or  $Eff_{pv} < Q_1 - 1.5IQR$  [25]. Here,  $Eff_{pv}$  denotes the PV efficiency of the historical data.  $Q_1$  and  $Q_3$  denote the first and third quantiles of all values of  $Eff_{pv}$ .  $IQR$  denotes the interquartile range (i.e.,  $Q_3 - Q_1$ ).

Step (2) Analog approach for hour-ahead power forecast: As aforementioned, the previous reported analog approaches select the training data collected at the historical period of the NWP predictions similar to the target forecast time. In this paper, the proposed analog method applies the meteorological observations of the hour before the analyzed historical hour and the target hour as the one-hour-ahead meteorological forecasts instead of the NWP forecasts. This approach is motivated by the one-hour-ahead persistence forecast method, which assumes that the hour-ahead condition will be the same as the present condition. Accordingly, the meteorological observations at the hour before can be roughly considered as the hour-ahead meteorological predictions, and thus be used to select similar historical data.

Step (3) NNE forecast method: For ensemble forecasting, diversity is a key feature, as there will be less improvement if the base predictors return similar decisions [26]. In this paper, the diversity of the base predictors is generated by changing the numbers of hidden layer neurons of the neural networks (NNs). Specifically, within an NNE framework, the base predictors are created and trained using the training data selected in Step (2) for the target hour; then the forecast results of the base predictors are blended to derive the final predictions.

### 3.2. Analog Approach for Hour-Ahead Power Forecasting

Analog methods rely on the fact that the past may reappear in the future when the external conditions are the same. Therefore, it firstly needs to measure the similarity of the past to the current [15]. The weighted distance is the most commonly used metric [11,13]. In one study [15], a Taylor expansion for the distance metric is used. In other studies [5,11], both the astronomical and the meteorological features are used to select analogs for training purposes. In addition, the analog method can be used to generate both deterministic and probabilistic forecasts [11,13,14]. It assumes that if a forecast made in the past is analogous to a current meteorological forecast, then it is likely to produce the same error characteristics as is probable in a current forecast.

Motivated by the widely used analog methods based on the NWP data, we propose a new analog approach which is suitable for one-hour-ahead PV power forecasting, without using the NWP data. In the NWP-based analog approaches, the similarity between the past and the present is measured using the meteorological forecasts of NWP models. In this paper, we use the persistence model for one-hour-ahead meteorological forecasts. The persistence model utilizes the most recent measurements as the forecast values. The persistence method is used here since it has high accuracies for short timeframe forecasts (e.g., one-hour-ahead forecast) [5], and its simplicity can reduce the whole computational complexity significantly. Furthermore, the astronomical information is also applied as critical physical factors for the proposed analog approach.

The steps of the proposed analog approach for one-hour-ahead PV power forecasting are shown in Figure 4. The analog training data for the target forecast time ( $t_g$ ) are chosen from the training data candidates tested at  $t_i$  ( $i = 1, \dots, n_{train}$ ).

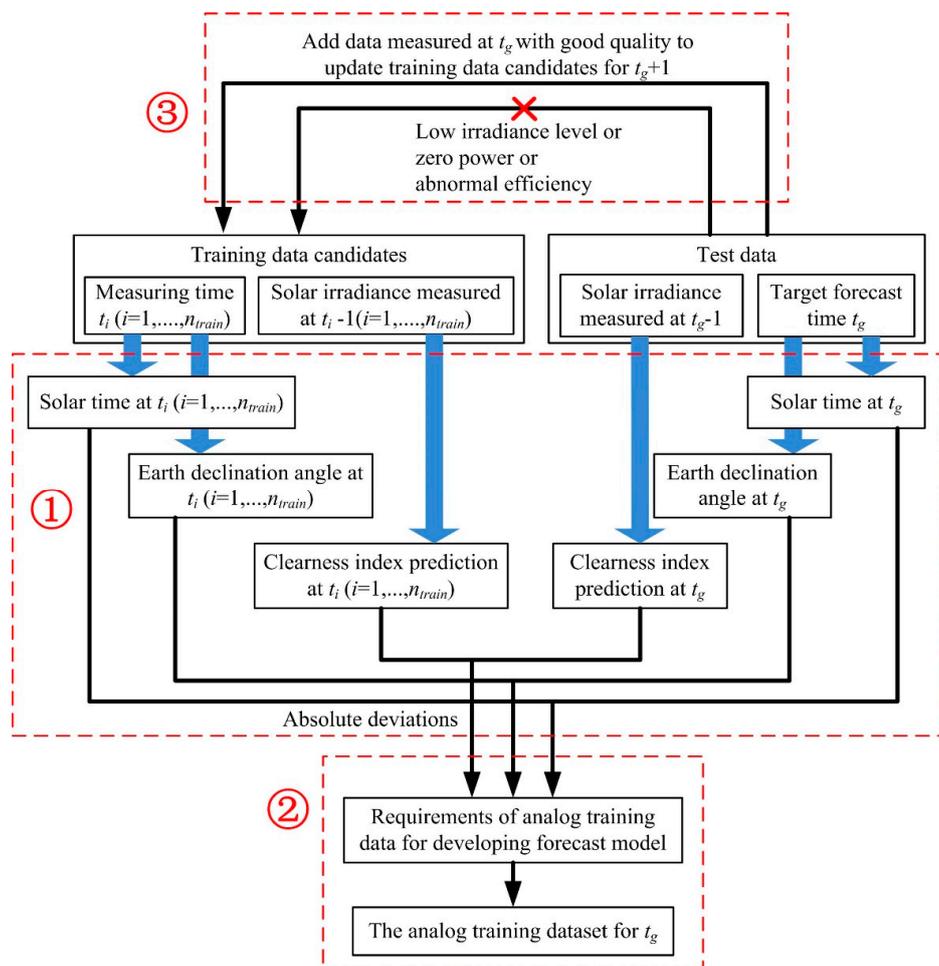


Figure 4. Steps of the proposed analog approach.

Firstly, both astronomical information including solar time ( $t_{solar}$ ) and earth declination angle ( $\delta$ ), and meteorological feature (i.e., the one-hour-ahead prediction of the clearness index ( $K_T$ )) are utilized as the features to measure the similarity degree of the weather conditions. In a study by Zhang and colleagues [5], solar time and earth declination angle are applied as physical factors. On the other hand, weather conditions have notable effects on PV performance, and have been widely taken into consideration for PV power forecasting [27,28]. The clearness index is defined as the ratio of the horizontal global irradiance passing through the aerosphere to the Earth’s surface to the extraterrestrial solar radiation [27]. Its value reflects the variations of cloud cover and atmospheric conditions, which are the causes of the PV power variability. Therefore, we apply the prediction of the hourly  $K_T$  at  $t_i$  ( $K_{T\_fore}(t_i)$ ) predicted by the persistence model as the meteorological feature, which can be calculated using Equations (1) and (2).

$$GHI_{fore}(t_i) = GHI_{meas}(t_i - 1), \quad i = 1, \dots, n \tag{1}$$

$$K_{T\_fore}(t_i) = \frac{GHI_{fore}(t_i)}{GHI_{ext}(t_i)}, \quad i = 1, \dots, n \tag{2}$$

where  $GHI_{fore}(t_i)$ ,  $GHI_{meas}(t_i - 1)$ , and  $GHI_{ext}(t_i)$  denote the forecasted horizontal global irradiance on the Earth’s surface at  $t_i$ , the measured horizontal global irradiance at  $t_i - 1$ , and the extraterrestrial solar radiation at  $t_i$ .

Secondly, the absolute deviations between the values of the three features at  $t_i$  and  $t_g$  are compared with the corresponding limits as shown in Equations (3)–(5). When the three absolute deviations satisfy the requirements at the same time, the data measured at  $t_i$  are considered as analog data of  $t_g$  and

selected to constitute the training data for  $t_g$ . The limits of the absolute deviations between the values of  $t_{solar}$ ,  $\delta$ , and  $K_{T\_fore}$  at  $t_i$  and  $t_g$  (i.e.,  $L_t$ ,  $L_\delta$  and  $L_{KT}$ ) are defined by trial and error. The definition of these parameters is discussed in Section 4.2.

$$|t_{solar}(t_g) - t_{solar}(t_i)| \leq L_t, \quad i = 1, \dots, n \quad (3)$$

$$|\delta(t_g) - \delta(t_i)| \leq L_\delta, \quad i = 1, \dots, n \quad (4)$$

$$|K_{T\_fore}(t_g) - K_{T\_fore}(t_i)| \leq L_{KT}, \quad i = 1, \dots, n \quad (5)$$

The final step is to update the training data candidates for the next forecast time (i.e.,  $t_g + 1$ ). When the measurements at  $t_g$  are available, their qualities are checked. The quality testing procedure is the same as the data preprocessing in Section 3.1. The data with low solar irradiance, zero power, and abnormal efficiency cannot be added to the training data candidates. It can be seen that updating the training data candidates can increase the size of the training data, and also take into account the latest effects (e.g., aging of the PV modules).

### 3.3. Neural Network Ensemble (NNE) Forecast Method

A great number of the existing research literature on ensemble PV power forecasting methods blends forecast results from different NWP models to get the final forecasted PV power. In a study by Zhang and colleagues [5], a day-ahead hourly PV power forecasting method is proposed, which blends forecasted power obtained by three NWP models using the weighting coefficients. For each forecasting hour, one set of weights is obtained by minimizing the forecast normalized mean absolute error (NMAE). The strategy to compute the weights of multiple forecasts from NWP models is to minimize the continuous ranked probability score [20]. Moreover, using the mean value of all forecasts as the final forecast is the simplest way to combine forecasts. It often outperforms more complex methods [19]. However, the NWP-based ensemble forecast methods are relatively less accurate for horizons shorter than six hours.

Another category of ensemble methods applies the ensembles of neural networks (NNs) for PV power forecasting. In a study by Raza and co-workers [22], feed forward neural network (FNN), Elman backpropagation network, and cascade-forward backpropagation networks are trained as base predictors, and then aggregated using Bayesian model averaging. Except for training different types of ANNs, other strategies to build ANN predictors of different structures include changing the number of hidden layer neurons and using diverse training data [29,30]. In this paper, we investigate the application of ensembles of NNs for the one-hour ahead PV power forecasting. Since the base learner used in many ensemble methods is a weak learner and the neural network is a strong learner, using neural networks as a base learner can improve the model performance.

The proposed NNE framework is a systematic ensemble of the FNN predictors, as shown in Figure 5. Several common ANN models, including the FNN, the radial basis function (RBF) network, and the Elman network, are compared. The FNN is chosen as the base predictor for its good performance. The FNN has a single hidden layer and is trained using the backpropagation (BP) algorithm and genetic algorithm (GA) for better network learning. The PV power predictions at  $t_g$  are denoted by  $F_{base}(t_g)$ , which are the outputs of the base predictors; the global horizontal irradiance ( $GHI_{meas}$ ), the ambient temperature ( $T_{amb}$ ), the wind speed ( $ws$ ), and the PV power ( $P$ ) measured at the prior hour before  $t_g$  (denoted by  $GHI_{meas}(t_g - 1)$ ,  $T_{amb}(t_g - 1)$ ,  $ws(t_g - 1)$  and  $P(t_g - 1)$ , respectively) act as the inputs of the base predictors.

The construction procedure of the proposed NNE framework is shown in Figure 6. The NNE framework is developed based on the training data selected by the analog approach in Section 3.2. There are 9 base predictors in the proposed NNE framework. In the NNE framework, the inputs are  $GHI(t_g - 1)$ ,  $T_{amb}(t_g - 1)$ ,  $ws(t_g - 1)$ , and  $P(t_g - 1)$ . Every base FNN model has 4 input neurons as shown in Figure 5. In order to build diverse base predictors, the numbers of hidden layer neurons are different for each FNN model. We determine the number of hidden layer neurons based on a rule of

thumb [31]. In the first FNN predictor (Network1), 4 hidden layer neurons are selected, the second FNN model contains 1 more than the first one, which is 5, and so on. The last or ninth FNN model contains 12 hidden layer neurons. The computational complexity and forecast accuracy of each FNN model vary due to the network architecture and the number of neurons in the hidden layer.

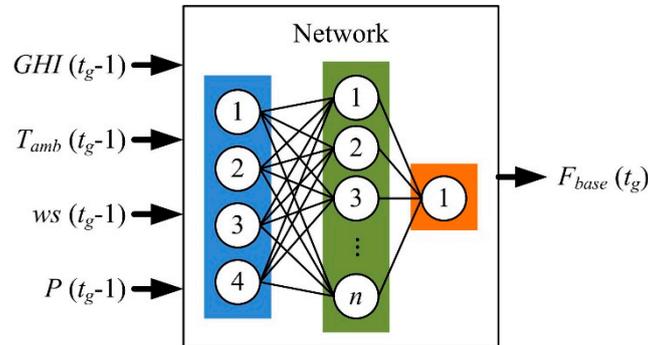


Figure 5. The structure of base neural network (NN) predictor.

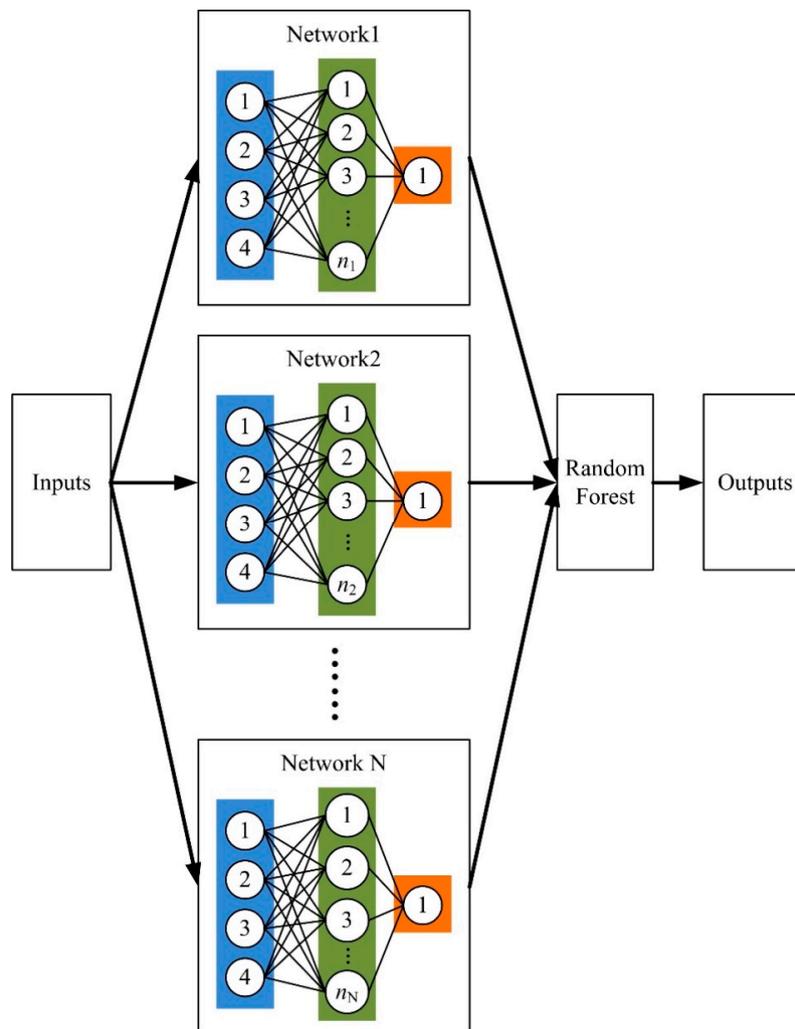


Figure 6. Construction of neural network ensemble (NNE).

We use the RF algorithm as the ensemble learning tool to combine the forecast results of the base predictors. The RF algorithm is actually an ensemble of decision trees (DTs). It employs the bootstrap

samples from the training data, and the random node and the split point selection to grow the ensemble of trees [32]. In this paper, the parameters of the RF are selected empirically: the number of DTs is 200, the number of samples to select at random for each decision split is 3 (i.e., 9/3), and the minimum number of samples per leaf is 5. The ensemble result for the forecast time  $t_g$  is made by averaging the individual tree's outputs. As the training data is updated every hour, the NNE framework should be reconstructed every hour as well using the updated training data.

#### 4. Numerical Results

In this section, the performance of the proposed methodology for hour-ahead PV power forecasting is tested using the measurements of a randomly chosen sub-array in a study by Wang and colleagues [23]. In addition, in order to demonstrate the effectiveness of the proposed method, several benchmark methods for one-hour-ahead PV power forecasting are applied for comparison.

Normalized root mean square error (NRMSE) and normalized mean absolute error (NMAE) are widely applied to evaluate the accuracy of the forecasting methods. They are formulated as:

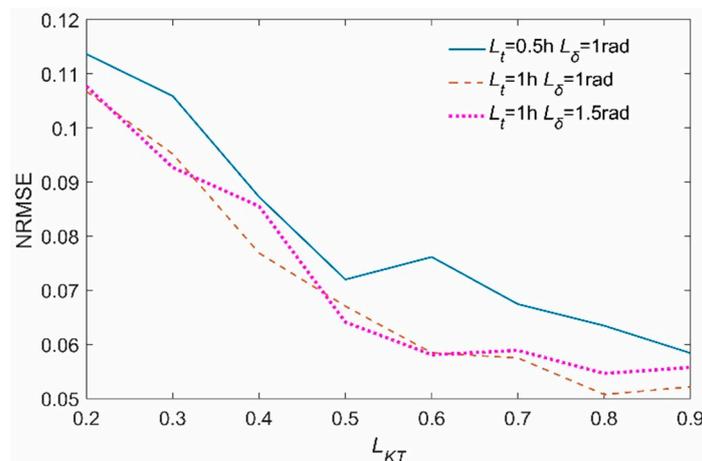
$$NRMSE = \frac{\sqrt{\frac{1}{n_{test}} \sum_{i=1}^{n_{test}} (F(t_i) - P(t_i))^2}}{Capacity} \times 100\% \quad (6)$$

$$NMAE = \frac{\frac{1}{n_{test}} \sum_{i=1}^{n_{test}} |F(t_i) - P(t_i)|}{Capacity} \times 100\% \quad (7)$$

where  $n_{test}$  is the number of the test data;  $F$ ,  $P$ , and  $Capacity$  denote the predicted, the measured, and the rated output power of the tested PV sub-array, respectively.

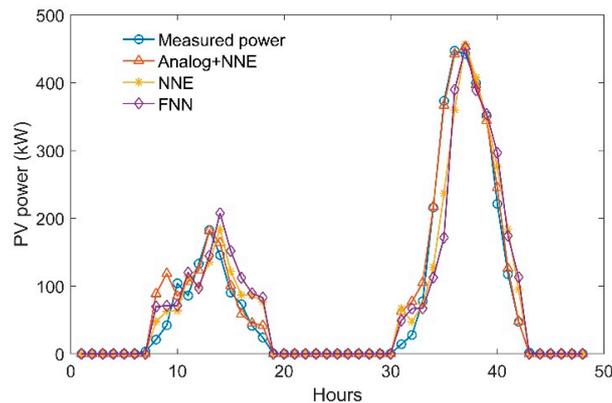
##### 4.1. Forecast Performance of the Proposed Methods

Before developing the methodology for PV power forecast, the limits of the changes of  $t_{solar}$ ,  $\delta$ , and  $K_T$  are set for the proposed analog approach. The target of the analog approach is to select the training data which are tested at the similar conditions from the historical observations, while guaranteeing enough data for the development of models. To this end, the limits (i.e.,  $L_t$ ,  $L_\delta$ , and  $L_{KT}$ ) are chosen for minimizing the NRMSE obtained by the analog plus NNE method tested on the validation data. The rough changing trend of the values of the NRMSE with the varied  $L_t$ ,  $L_\delta$ , and  $L_{KT}$  is shown in Figure 7. It can be seen that the minimized NRMSE occurs at  $L_t$  of 1 h,  $L_\delta$  of 1 rad, and  $L_{KT}$  of 0.8, respectively. Accordingly,  $L_t$ ,  $L_\delta$ , and  $L_{KT}$  are set to be 1 h, 1 rad, and 0.8.



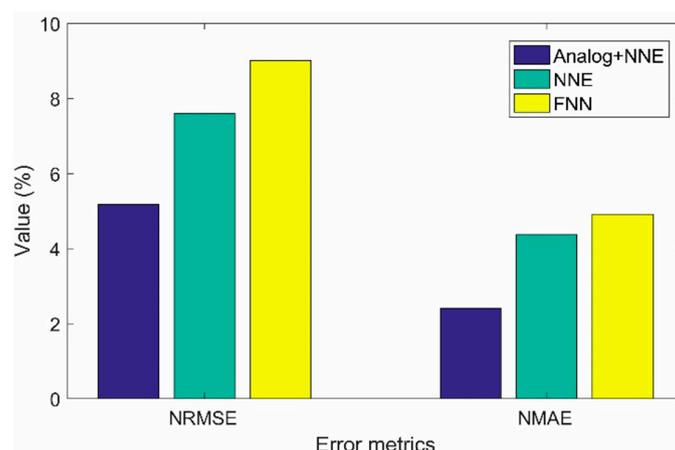
**Figure 7.** Values of normalized root mean square error (NRMSE) by using the different limits of the changes of  $t_{solar}$ ,  $\delta$ , and  $K_T$ .

For instance, the forecasting results and the measured power from 14 January 2016 00:00 to 16 January 2016 00:00 are shown in Figure 8. Different methods, including the proposed analog plus NNE (analog + NNE) method, the NNE method, and the single FNN model, are compared in this period. In Figure 8, it can be seen that the proposed analog + NNE method has forecasting results closest to the measured power. On the other hand, the NNE method outperforms the single FNN model at most test points. It proves that the ensemble technique based on the RF algorithm is effective to enhance the forecast accuracy of the single FNN model. Since the analog + NNE method uses the same forecast model with the NNE method, the improvements are made by the proposed analog approach selecting training data.



**Figure 8.** Forecast comparison of the methods from 14 January 2016 00:00 to 16 January 2016 00:00.

The values of the *NRMSE* and *NMAE* computed for the methods are shown in Figure 9. The proposed method combining the analog approach for one-hour-ahead forecast with the NNE provides the most accurate results. Compared to the single FNN model, the NNE approach reduces the forecast error metrics from 9.02% *NRMSE* and 4.92% *NMAE* to 7.60% *NRMSE* and 4.38% *NMAE*. This demonstrates the effectiveness of the proposed ensemble method based on the RF. The comparison between the analog + NNE and the NNE methods indicates that the proposed analog approach can reduce the error from 7.60% *NRMSE* and 4.38% *NMAE* to 5.18% *NRMSE* and 2.42% *NMAE*. This shows the importance of selecting the proper data for the model training.



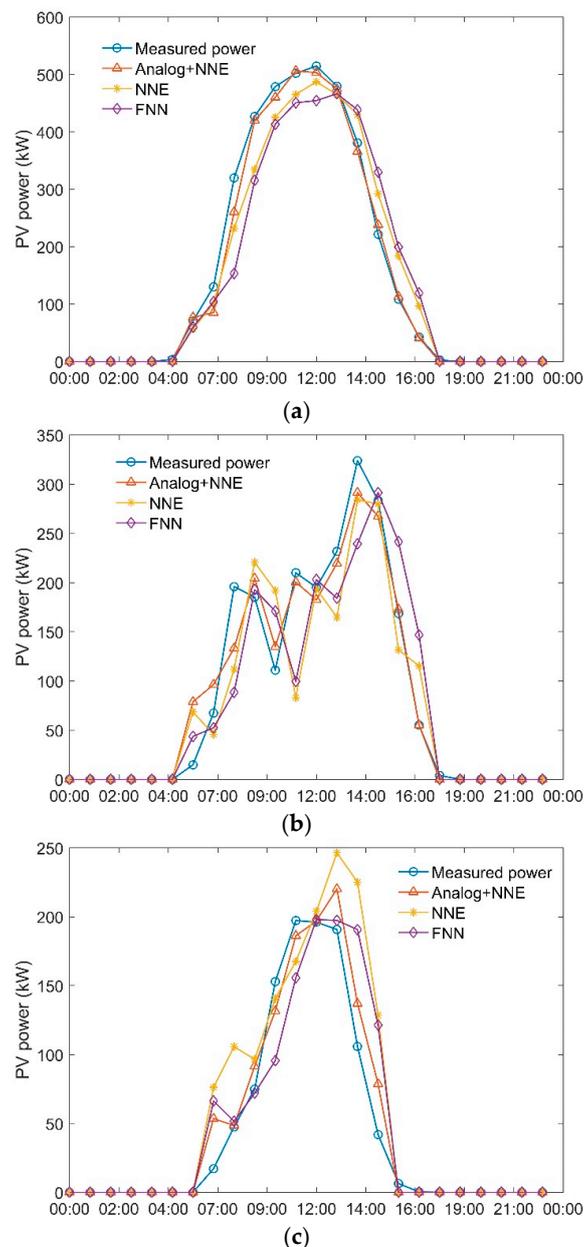
**Figure 9.** Error metrics comparison of the methods.

#### 4.2. Impact of Weather Types on Forecasting

For PV power forecasting, weather type is widely taken into account. For instance, in a study by Yang and co-workers [28], features were extracted from the collected historical data and then used to classify the weather type. Sub-models were further established for the corresponding weather

conditions. Generally, solar irradiance is stable on sunny days, however, it fluctuates on partial cloudy days. Influenced by varying solar irradiation, PV power output is intermittent and undispachable when solar irradiance is varying. In addition, the forecast accuracies change with the different weather types [33]. In this section, the forecast performances of the methods under different weather types are analyzed.

The measured and the predicted PV power under the different weather types are shown in Figure 10. The days can be divided into different weather types based on the value of the daily  $K_T$  [30]. The clear day has a daily  $K_T$  more than 0.45; the partially cloudy day corresponds to a daily  $K_T$  in the range of (0.25, 0.45); and the cloudy day has a daily  $K_T$  less than 0.25. It can be seen that the PV power highly fluctuates due to the moving clouds on a partially cloudy day as shown in Figure 10b. Due to the shadings of clouds, the power generations on a cloudy day in Figure 10c are much smaller than that on a clear day as shown in Figure 10a.



**Figure 10.** Examples of forecast results at different weather conditions: (a) clear day (daily  $K_T = 0.65$ ); (b) partially cloudy day (daily  $K_T = 0.44$ ); (c) cloudy day (daily  $K_T = 0.21$ ).

Table 2 lists the error metrics computed for the forecasting methods under the three weather types. The proposed analog + NNE method outperforms the comparative methods in terms of accuracy under all weather conditions. During the clear days, the analog plus NNE method reduces the *NRMSE* and *NMAE* by 52.81% and 62.00% compared with the single FNN method. The proposed analog approach reduces *NRMSE* and *NMAE* by 50.77% and 59.96%. This is derived by comparing the analog plus NNE method with the NNE method. The NNE method does better than the single FNN method. It reduces *NRMSE* and *NMAE* of the single FNN method by 4.16% and 5.1%, respectively. It can be inferred that the analog approach makes a greater contribution for improving the forecasting accuracy than the ensemble technique based on the RF.

**Table 2.** Forecast performances of the methods under different weather types.

| Error Metrics | Methods      | Weather Types                      |   |                                     |
|---------------|--------------|------------------------------------|---|-------------------------------------|
|               |              | Clear Day<br>(Daily $K_T > 0.45$ ) | Partially Cloudy Day<br>( $0.25 < \text{Daily } K_T < 0.45$ ) | Cloudy Day<br>(Daily $K_T < 0.25$ ) |
| <i>NRMSE</i>  | Analog + NNE | 3.86%                              | 6.02%   | 5.74%                               |
|               | NNE          | 7.84%                              | 9.69%   | 6.45%                               |
|               | FNN          | 8.18%                              | 9.92%   | 6.79%                               |
| <i>NMAE</i>   | Analog + NNE | 1.79%                              | 2.93%   | 2.74%                               |
|               | NNE          | 4.47%                              | 5.23%   | 3.49%                               |
|               | FNN          | 4.71%                              | 5.42%   | 3.73%                               |

$K_T$ : Clearness index.

In Table 2, all the methods have the highest error during the partially cloudy days. On a partially cloudy day, the observed PV power fluctuates significantly due to the fluctuating solar irradiance. Thus, we deduce that it is difficult for the predicted PV power to track the rapidly changing trends of the measured PV power accurately.

Furthermore, there is another noteworthy point in Table 2 to discuss. The analog plus NNE method produces the most accurate results during clear days. This coincides with the related results in literature (e.g., [30]). However, in Table 2, the NNE method and the single FNN method give lower *NRMSE* and *NMAE* values during the cloudy days than the clear days. As shown in Figure 10, due to the shading effects of clouds, the PV power productions on a cloudy day are approximately less than half the power generations on a clear day. This indicates that the absolute deviation between the predicted and the measured PV power (i.e.,  $F(t_i) - P(t_i)$ ) in Equations (6) and (7) on a clear day will be essentially larger than that on a cloudy day. Thus, the larger *NRMSE* and *NMAE* of the NNE and the FNN methods during the clear days are mainly caused by the larger absolute deviation between  $F(t_i)$  and  $P(t_i)$  compared with the cloudy days.

#### 4.3. Comparison Results with Benchmark Methods

In this section, the proposed analog plus NNE method is compared with six benchmark methods for PV power forecasting, namely, the persistence model, the support vector regression (SVR) model, the linear regression model, the RF model, the gradient boosting model, and XGBoost model.

##### A. Persistence model

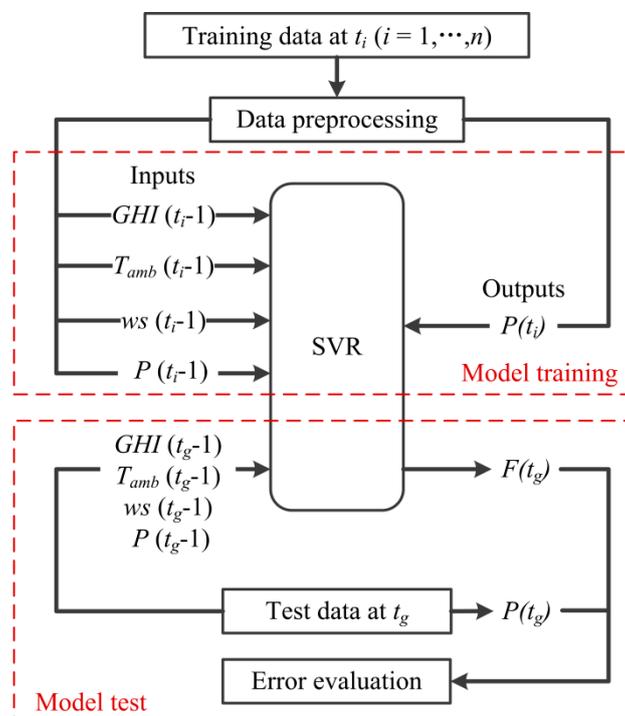
The first benchmark model is the persistence model, which is typically used for short-term forecasting, such as one-hour-ahead forecasting [5]. The persistence model is based on the memory properties of time series. It assumes that the conditions at the target forecast time are the same as that at the most recent moment. Accordingly, the forecast value at  $t_g$  (i.e.,  $F(t_g)$ ) obtained by the persistence model is equal to the last measured value (i.e.,  $P(t_g - 1)$ ):

$$F(t_g) = P(t_g - 1) \quad (8)$$

Despite its simplicity, the persistence model can achieve tolerable results for short-term PV power forecasting. Therefore, it widely acts as a benchmark method to test the qualities of new forecasting approaches [5]. In this section, the persistence model is tested on the same test data.

### B. SVR model

The SVR model has been widely used to predict PV power with good generalization capability [5]. The outstanding aspect of SVR is that it realizes the best trade-off between the empirical error and the model complexity [34]. SVR is able to perform an improved non-linear regression by the kernel trick and other optimization features. In this work, an RBF kernel is applied. The parameters of the SVR model are optimized by the grid-search with cross-validation. In order to make comparison with the proposed method, we use the same inputs, training data, and data preprocessing approach for the development of the SVR model. The performance of the SVR model is tested on the same test data. This procedure is shown in Figure 11.



**Figure 11.** The framework of support vector regression (SVR) benchmark model.

### C. Linear regression model

In addition to the above two models, the linear regression model is also used to compare with the proposed analog plus NNE method as the base line of the evaluation of the forecast results.

### D. RF model

RF model has been widely used to predict PV power output. The random forest is an ensemble method that aggregates the output of several uncorrelated decision trees [35]. In this study, the RF model is tested on the same test data.

### E. Gradient boosting model

Gradient boosting is a machine learning technique for regression problems which is typically used for short-term forecasting. In this study, a gradient boosting model is tested on the same test data.

## F. XGBoost model

In addition to the above five models, XGBoost model is also used to compare with the proposed analog plus NNE method. XGBoost has shown superiorities in many power forecasting projects in recent years due to its advantages [36]. The performance of the XGBoost model is tested on the same test data.

Forecast skill (FS) metric is a popular metric to evaluate the forecasting accuracy improved by a method compared with the benchmark methods [9]. It is computed by:

$$FS = 1 - \frac{error}{error_b} \quad (9)$$

where  $error$  and  $error_b$  refer to the error values of the proposed method and the benchmark models, respectively. The evaluated method with more accurate forecasts yields a positive skill. The larger the FS, the better the performance of the method.

Table 3 highlights that the proposed analog plus NNE method produces a higher forecast accuracy in comparison with the persistence model, the SVR model, the linear regression model, the RF model, the gradient boosting model, and XGBoost model. The linear regression model has the largest value for the *NRMSE* error metric, and the gradient boosting model has the largest value for the *NMAE* error metric. Compared with the other five methods, the SVR model gives a better forecasting result in terms of accuracy. However, the improvement made by the SVR model is relatively low (FS of *NRMSE* of 0.03, and FS of *NMAE* of 0.01). This indicates that, compared with the other five methods, the advantage of the SVR model is not obvious for one-hour-ahead PV power forecasting. The analog plus NNE method can significantly reduce the forecast error. As shown in Table 3, compared with the persistence method, the *NRMSE* and *NMAE* are reduced by 43.51% and 50.71% due to the proposed method. The *NRMSE* and *NMAE* of the SVR model are reduced by 41.80% and 50.31%. The *NRMSE* and *NMAE* of the linear regression model are reduced by 52.13% and 51.41%. The *NRMSE* and *NMAE* of the RF model are reduced by 42.38% and 51.70%. The *NRMSE* and *NMAE* of the gradient boosting model are reduced by 50.76% and 59.33%. The *NRMSE* and *NMAE* of the XGBoost model are reduced by 46.43% and 53.46%.

**Table 3.** Forecast performance comparison of the proposed method with the benchmarks.

| Methods                         | Error Metrics | FS     | Error Reduction by Analog + NNE Method |        |
|---------------------------------|---------------|--------|--|--------|
| Persistence                     | <i>NRMSE</i>  | 9.17%  | 0                                      | 43.51% |
|                                 | <i>NMAE</i>   | 4.91%  | 0                                      | 50.71% |
| Support vector regression (SVR) | <i>NRMSE</i>  | 8.90%  | 0.03                                   | 41.80% |
|                                 | <i>NMAE</i>   | 4.87%  | 0.01                                   | 50.31% |
| Linear regression               | <i>NRMSE</i>  | 10.82% | −0.18                                  | 52.13% |
|                                 | <i>NMAE</i>   | 4.98%  | −0.01                                  | 51.41% |
| Random forest (RF) model        | <i>NRMSE</i>  | 8.99%  | 0.02                                   | 42.38% |
|                                 | <i>NMAE</i>   | 5.01%  | −0.02                                  | 51.70% |
| Gradient boosting               | <i>NRMSE</i>  | 10.52% | −0.15                                  | 50.76% |
|                                 | <i>NMAE</i>   | 5.95%  | −0.21                                  | 59.33% |
| XGBoost                         | <i>NRMSE</i>  | 9.67%  | −0.05                                  | 46.43% |
|                                 | <i>NMAE</i>   | 5.20%  | −0.06                                  | 53.46% |
| Analog + NNE                    | <i>NRMSE</i>  | 5.18%  | 0.44                                   |        |
|                                 | <i>NMAE</i>   | 2.42%  | 0.51                                   |        |

Additionally, in this study, the proposed analog plus NNE method, persistence model, SVR model, linear regression model, and RF model are implemented using the Matlab script. The gradient boosting model and XGBoost model are implemented using the Python script.

## 5. Conclusions

This paper proposes an analog plus NNE method for one-hour-ahead PV power forecasting. The analog approach utilizes certain metrics to describe how much the test conditions of the target

forecast hour is similar to that of the historical hours. In this work, solar time and earth declination angle at the target forecast hour, and the one-hour-ahead clearness index prediction of the target forecast hour act as the metrics. The analog method selects the training data which are tested at conditions similar to that of the target forecast hour by setting limits for the metrics. Based on the selected data, an NNE method combining multiple FNN predictors of different numbers of hidden layer neurons by RF is applied to train the forecast model at the target forecast hour. The methodology is tested using the measurements of a sub-array in a grid-connected PV plant. Results show that the proposed method can significantly improve the forecast accuracy compared with the benchmark methods. The limitation of the proposed approach is that the forecast horizon considered for the very-short-term PV power forecasting is one-hour-ahead only, which is objectively limited by the current experimental conditions. In the future research work, the proposed method can be further extended to apply to longer term PV power forecasting, such as one-day-ahead.

**Author Contributions:** Conceptualization, J.W. (Jingyue Wang), Z.Q., J.W. (Jingyi Wang), and Y.P.; data curation, Y.P.; formal analysis, J.W. (Jingyue Wang) and J.W. (Jingyi Wang); funding acquisition, Z.Q.; methodology, J.W. (Jingyue Wang), Z.Q., J.W. (Jingyi Wang), and Y.P.; software, J.W. (Jingyue Wang) and J.W. (Jingyi Wang); supervision, Z.Q.; writing—original draft, J.W. (Jingyue Wang); writing—review and editing, Z.Q., J.W. (Jingyi Wang), and Y.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Nature Science Foundation of China (grant number 61573046) and the Program for Changjiang Scholars and Innovative Research Team in University (grant number IRT1203).

**Conflicts of Interest:** The authors declare no conflict of interest.

## Nomenclature

|                              |   |
|------------------------------|---|
| $Eff_{pv}$                   | PV efficiency   |
| $Q_1$ and $Q_3$              | First and third quantiles of $Eff_{pv}$   |
| IQR                          | Interquartile range (i.e., $Q_3 - Q_1$ )  |
| $t_g$                        | Target forecast time  |
| $t_i$                        | Training data candidates test time  |
| $t_{solar}$                  | Solar time  |
| $\delta$                     | Earth declination angle   |
| $K_T$                        | Clearness index   |
| $K_{T\_fore}$                | One hour-ahead forecast of the hourly $K_T$   |
| $GHI_{fore}$                 | Forecasted horizontal global irradiance on the earth's surface  |
| $GHI_{meas}$                 | Measured horizontal global irradiance   |
| $GHI_{ext}$                  | Extraterrestrial solar radiation  |
| $L_t, L_\delta$ and $L_{KT}$ | Absolute deviations between the values of $t_{solar}$ , $\delta$ and $K_{T\_fore}$ at $t_i$ and $t_g$ |
| $F_{base}$                   | PV power predicted by the base predictors   |
| $T_{amb}$                    | Ambient temperature   |
| $ws$                         | Wind speed  |
| $P$                          | Measured PV power   |
| $n_{test}$                   | Number of the test data   |
| Capacity                     | Rated output power of PV array  |
| error and $error_b$          | Error values of the proposed method and the benchmark models  |
| PV                           | Photovoltaic  |
| NWP                          | Numerical weather prediction  |
| ANN                          | Artificial neural network   |
| NNE                          | Neural network ensemble   |
| RF                           | Random forest   |
| NN                           | Neural network  |
| NMAE                         | Normalized mean absolute error  |

|       |                                   |
|-------|-----------------------------------|
| FNN   | Feed forward neural network       |
| BP    | Backpropagation                   |
| GA    | Genetic algorithm                 |
| DT    | Decision tree                     |
| NRMSE | Normalized root mean square error |
| SVR   | Support vector regression         |
| RBF   | Radial basis function             |
| FS    | Forecast skill                    |

## References

1. Das, U.K.; Tey, K.S.; Seyedmahmoudian, M.; Mekhilef, S.; Idris, M.Y.I.; Van Deventer, W.; Horan, B.; Stojcevski, A. Forecasting of photovoltaic power generation and model optimization: A review. *Renew. Sustain. Energy Rev.* **2018**, *81*, 912–928. [[CrossRef](#)]
2. Sobri, S.; Koochi-Kamali, S.; Rahim, N.A. Solar photovoltaic generation forecasting methods: A review. *Energy Convers. Manag.* **2018**, *156*, 459–497. [[CrossRef](#)]
3. Wan, C.; Zhao, J.; Song, Y.; Xu, Z.; Lin, J.; Hu, Z. Photovoltaic and solar power forecasting for smart grid energy management. *CSEE J. Power Energy Syst.* **2015**, *1*, 38–46. [[CrossRef](#)]
4. Gigoni, L.; Betti, A.; Crisostomi, E.; Franco, A.; Tucci, M.; Bizzarri, F.; Mucci, D. Day-Ahead Hourly Forecasting of Power Generation from Photovoltaic Plants. *IEEE Trans. Sustain. Energy* **2019**, *9*, 831–842. [[CrossRef](#)]
5. Zhang, X.; Li, Y.; Lu, S.; Hamann, H.; Hodge, B.-M.; Lehman, B. A Solar Time Based Analog Ensemble Method for Regional Solar Power Forecasting. *IEEE Trans. Sustain. Energy* **2018**, *10*, 268–279. [[CrossRef](#)]
6. Huang, R.; Huang, T.; Gadh, R.; Li, N. Solar generation prediction using the ARMA model in a laboratory-level micro-grid. In Proceedings of the IEEE 3rd International Conference on Smart Grid Communications (SmartGridComm), Tainan, Taiwan, 5–8 November 2012; pp. 528–533.
7. Jiang, H.; Dong, Y. Forecast of hourly global horizontal irradiance based on structured Kernel Support Vector Machine: A case study of Tibet area in China. *Energy Convers. Manag.* **2017**, *142*, 307–321. [[CrossRef](#)]
8. İzgi, E.; Öztopal, A.; Yerli, B.; Kaymak, M.K.; Şahin, A.D. Short–mid-term solar power prediction by using artificial neural networks. *Sol. Energy* **2012**, *86*, 725–733. [[CrossRef](#)]
9. Yang, D.; Kleissl, J.; Gueymard, C.A.; Pedro, H.T.C.; Coimbra, C.F.M. History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. *Sol. Energy* **2018**, *168*, 60–101. [[CrossRef](#)]
10. Yang, D.; Dong, Z. Operational photovoltaics power forecasting using seasonal time series ensemble. *Sol. Energy* **2018**, *166*, 529–541. [[CrossRef](#)]
11. Alessandrini, S.; Delle Monache, L.; Sperati, S.; Cervone, G. An analog ensemble for short-term probabilistic solar power forecast. *Appl. Energy* **2015**, *157*, 95–110. [[CrossRef](#)]
12. Cervone, G.; Clemente-Harding, L.; Alessandrini, S.; Delle Monache, L. Short-term photovoltaic power forecasting using Artificial Neural Networks and an Analog Ensemble. *Renew. Energy* **2017**, *108*, 274–286. [[CrossRef](#)]
13. Delle Monache, L.; Eckel, F.; Rife, D.; Nagarajan, B.; Searight, K. Probabilistic Weather Prediction with an Analog Ensemble. *Mon. Weather Rev.* **2013**, *141*, 3498–3516. [[CrossRef](#)]
14. Haupt, S.; Kosovic, B. Variable Generation Power Forecasting as a Big Data Problem. *IEEE Trans. Sustain. Energy* **2016**, *8*, 725–732. [[CrossRef](#)]
15. Akyurek, B.O.; Akyurek, A.S.; Kleissl, J.; Rosing, T.Š. TESLA: Taylor expanded solar analog forecasting. In Proceedings of the IEEE International Conference on Smart Grid Communications (SmartGridComm), Venice, Italy, 3–6 November 2014; pp. 127–132. [[CrossRef](#)]
16. Zorita, E.; Von Storch, H. The Analog Method as a Simple Statistical Downscaling Technique: Comparison with More Complicated Methods. *J. Clim.* **1999**, *12*, 2474–2489. [[CrossRef](#)]
17. Abuella, M.; Chowdhury, B. Random Forest Ensemble of Support Vector Regression Models for Solar Power Forecasting. In Proceedings of the IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA, 23–26 April 2017; pp. 1–5.
18. Chowdhury, B. Improving combined solar power forecasts using estimated ramp rates: Data-driven post-processing approach. *IET Renew. Power Gener.* **2018**, *12*, 1127–1135.

19. Kourentzes, N.; Barrow, D.K.; Crone, S.F. Neural network ensemble operators for time series forecasting. *Expert Syst. Appl.* **2014**, *41*, 4235–4244. [[CrossRef](#)]
20. Thorey, J.; Chaussin, C.; Mallet, V. Ensemble forecast of photovoltaic power with online CRPS learning. *Int. J. Forecast.* **2018**, *34*, 762–773. [[CrossRef](#)]
21. Larson, D.P.; Nonnenmacher, L.; Coimbra, C.F.M. Day-ahead forecasting of solar power output from photovoltaic plants in the American Southwest. *Renew. Energy* **2016**, *91*, 11–20. [[CrossRef](#)]
22. Raza, M.Q.; Mithulananthan, N.; Summerfield, A. Solar output power forecast using an ensemble framework with neural predictors and Bayesian adaptive combination. *Sol. Energy* **2018**, *166*, 226–241. [[CrossRef](#)]
23. Wang, J.Y.; Qian, Z.; Zareipour, H.; Wood, D. Performance assessment of photovoltaic modules based on daily energy generation estimation. *Energy* **2018**, *165*, 1160–1172. [[CrossRef](#)]
24. Platon, R.; Martel, J.; Woodruff, N.; Chau, T.Y. Online Fault Detection in PV Systems. *IEEE Trans. Sustain. Energy* **2015**, *6*, 1200–1207. [[CrossRef](#)]
25. Mallor, F.; León, T.; De Boeck, L.; Van Gulck, S.; Meulders, M.; Van der Meerssche, B. A method for detecting malfunctions in PV solar panels based on electricity production monitoring. *Sol. Energy* **2017**, *153*, 51–63. [[CrossRef](#)]
26. Ren, Y.; Suganthan, P.N.; Srikanth, N. Ensemble methods for wind and solar power forecasting—A state-of-the-art review. *Renew. Sustain. Energy Rev.* **2015**, *50*, 82–91. [[CrossRef](#)]
27. Wang, F.; Zhen, Z.; Mi, Z.; Sun, H.; Su, S.; Yang, G. Solar irradiance feature extraction and support vector machines based weather status pattern recognition model for short-term photovoltaic power forecasting. *Energy Build.* **2015**, *86*, 427–438. [[CrossRef](#)]
28. Yang, H.; Huang, C.; Huang, Y.; Pai, Y. A Weather-Based Hybrid Method for 1-Day Ahead Hourly Forecasting of PV Power Output. *IEEE Trans. Sustain. Energy* **2014**, *5*, 917–926. [[CrossRef](#)]
29. Al-Dahidi, S.; Ayadi, O.; Alrbai, M.; Adeeb, J. Ensemble Approach of Optimized Artificial Neural Networks for Solar Photovoltaic Power Prediction. *IEEE Access* **2019**, *7*, 81741–81758. [[CrossRef](#)]
30. Raza, M.Q.; Mithulananthan, N.; Li, J.; Lee, K.Y.; Gooi, H.B. An Ensemble Framework for Day-Ahead Forecast of PV Output Power in Smart Grids. *IEEE Trans. Ind. Inform.* **2019**, *15*, 4624–4634. [[CrossRef](#)]
31. Gao, D.Q.; Zhu, H.J.; Nie, G.P. On the transformation mechanisms of multilayer perceptrons with sigmoid activation functions for classifications. In Proceedings of the International Joint Conference on Neural Networks, Portland, OR, USA, 20–24 July 2003; Volume 2, pp. 1173–1178. [[CrossRef](#)]
32. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
33. Shi, J.; Lee, W.-J.; Liu, Y.; Yang, Y.; Wang, P. Forecasting power output of photovoltaic system based on weather classification and support vector machine. In Proceedings of the IEEE Industry Applications Society Annual Meeting, Orlando, FL, USA, 9–13 October 2011; pp. 1–6.
34. Santamaría-Bonfil, G.; Reyes-Ballesteros, A.; Gershenson, C. Wind speed forecasting for wind farms: A method based on support vector regression. *Renew. Energy* **2016**, *85*, 790–809. [[CrossRef](#)]
35. Lahouar, A.; Mejri, A.; Slama, J.B.H. Importance based selection method for day-ahead photovoltaic power forecast using random forests. In Proceedings of the 2017 International Conference on Green Energy Conversion Systems (GECS), Hammamet, Tunisia, 23–25 March 2017; pp. 1–7. [[CrossRef](#)]
36. Dong, W.; Huang, Y.M.; Barry, L.; Ma, G.W. XGBoost algorithm-based prediction of concrete electrical resistivity for structural health monitoring. *Automat. Constr.* **2020**, *114*, 103155. [[CrossRef](#)]

