



## Article

# Time-Section Fusion Pattern Classification Based Day-Ahead Solar Irradiance Ensemble Forecasting Model Using Mutual Iterative Optimization

Fei Wang <sup>1,2,\*</sup> , Zhao Zhen <sup>1</sup>, Chun Liu <sup>3</sup>, Zengqiang Mi <sup>1</sup>, Miadreza Shafie-khah <sup>4</sup>  and João P. S. Catalão <sup>4,5,6,\*</sup>

- <sup>1</sup> State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Baoding 071003, China; zhenzhao@ncepu.edu.cn (Z.Z.); mizengqiang@ncepu.edu.cn (Z.M.)
- <sup>2</sup> Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
- <sup>3</sup> State Key Laboratory of Operation and Control of Renewable Energy & Storage Systems, China Electric Power Research Institute, Beijing 100192, China; liuchun@epri.sgcc.com.cn
- <sup>4</sup> C-MAST, University of Beira Interior, 6201-001 Covilhã, Portugal; miadreza@gmail.com
- <sup>5</sup> INESC TEC, Faculty of Engineering of the University of Porto, 4200-465 Porto, Portugal
- <sup>6</sup> INESC-ID, Instituto Superior Técnico, University of Lisbon, 1049-001 Lisbon, Portugal
- \* Correspondence: feiwang@ncepu.edu.cn (F.W.); catalao@ubi.pt (J.P.S.C.)

Received: 12 November 2017; Accepted: 8 January 2018; Published: 12 January 2018

**Abstract:** Accurate solar PV power forecasting can provide expected future PV output power so as to help the system operator to dispatch traditional power plants to maintain the balance between supply and demand sides. However, under non-stationary weather conditions, such as cloudy or partly cloudy days, the variability of solar irradiance makes the accurate PV power forecasting a very hard task. Ensemble forecasting based on multiple models established by different theory has been proved as an effective means on improving forecasting accuracy. Classification modeling according to different patterns could reduce the complexity and difficulty of intro-class data fitting so as to improve the forecasting accuracy as well. When combining the two above points and focusing on the different fusion pattern specifically in terms of hourly time dimension, a time-section fusion pattern classification based day-ahead solar irradiance ensemble forecasting model using mutual iterative optimization is proposed, which contains multiple forecasting models based on wavelet decomposition (WD), fusion pattern classification model, and fusion models corresponding to each fusion pattern. First, the solar irradiance is forecasted using WD based models at different WD level. Second, the fusion pattern classification recognition model is trained and then applied to recognize the different fusion pattern at each hourly time section. At last, the final forecasting result is obtained using the optimal fusion model corresponding to the data fusion pattern. In addition, a mutual iterative optimization framework for the pattern classification and data fusion models is also proposed to improve the model's performance. Simulations show that the mutual iterative optimization framework can effectively enhance the performance and coordination of pattern classification and data fusion models. The accuracy of the proposed solar irradiance day-ahead ensemble forecasting model is verified when compared with a standard Artificial Neural Network (ANN) forecasting model, five WD based models and a single ensemble forecasting model without time-section fusion classification.

**Keywords:** day-ahead forecasting; ensemble model; solar irradiance; data fusion; mutual iterative optimization

## 1. Introduction

In recent years, due to the shortages of fossil fuel and their adverse impacts on the environment, worldwide interest in the deployment of solar power generation is rapidly increasing [1].

In the year 2015, the solar PV market was up 25% over 2014 to a record 50 GW, lifting the global total to 227 GW. The annual PV market in 2015 was nearly 10 times the world's cumulative solar PV capacity of a decade earlier. China contributed significantly to global solar PV growth, the net PV capacity additions in 2015 and total PV capacity by the end-2015 of China all ranks first in the world [2].

However, the variable nature of renewable power generation is an obstacle to achieving a higher level of solar penetration into the power grid. With the increase in the amount of grid-connected PV plant and installed PV capacity, curtailment of solar generation started to become a serious challenge for China's solar PV sector [3].

The uncertainty and fluctuation of PV power output can be managed with several possible solutions, including increased demand-side participation, greater coordination to balance allocation among areas, more energy storage equipment, and larger power reserve capacity [4–7]. However, taking into account the economic and feasibility factors, PV power forecasting is still one of the most effective and economical ways to solve the uncertainty and fluctuation problems in PV system. Day-ahead PV power forecasting can provide the expected future PV power output of the next day 24 h, which is an important reference information for power generation planning and system scheduling [8]. Accurate solar forecasting is not only able to help system operators and planners to better manage the variability and uncertainty of PV power in advance, but also can benefit PV plant managers as they avoid possible penalties that are incurred due to deviations between forecasted and produced energy [9].

Solar irradiance is the main influence factor that affecting the power output of a PV plant, the accurate forecasting of irradiance is of great importance for step-wise PV power forecasting [10–13]. Generally, three kinds of method are applicable for solar irradiance forecasting at present, respectively image processing based method using total sky image or satellite image, numerical weather prediction (NWP) based method, and machine learning method using historical irradiance data. Image processing based methods are mainly applied in ultra-short term irradiance forecasting [14–16]. NWP data can be used for day-ahead irradiance forecasting [17], the performance of these NWP based methods mainly depends on the accuracy of NWP data. However, accurate NWP service is usually expensive and it may be unavailable in some remote areas, which restricted its application.

Therefore, in this paper, we focus on the machine learning based method that using historical data to forecast the future irradiance. Affected by the micro-meteorological environment of PV power plant, solar irradiance data sequence contains both stable, periodic component and fluctuant, random component. The former component varied over time every day appears more obvious on sunny days than overcast or rainy days, while the latter component is not that obvious because it is mainly affected by meteorological factors, such as cloud movement or showers. The proportion of the two different components also varies with different weather conditions. Therefore, forecasting models based on a single algorithm are not adequate to solve all the forecasting problems in different environments.

Ensemble model has proved to be an efficient means to solve this kind of forecasting problems with diversified data, as it can comprehensive advantages of a variety of algorithms and suit to multiple conditions [17–19]. However, how to utilize the forecasting results of each individual model in the total ensemble model is still a problem to be solved. Probabilistic forecasting approach is a common application of ensemble model as it can achieve the probability of different forecasting value, according to the outputs of multiple models [20,21]. In some situations, a unique result that can best provide the future information is still need, and a very common way is to calculate the average of the ensemble model outputs or use a mapping model to fuse the multiple results [17,22]. But, the fusion mapping relationship between forecasting results of different models and actual

irradiance is very complicated and changeable, which depends on the performance of each forecasting model, the meteorological environment around PV plant, the geographical environment of PV plant, the forecasting time, etc. Therefore, it is very difficult to distinguish and recognize the different fusion patterns by traditional artificial pattern recognition analysis. Additionally, once the environment of PV plant changes, the original model will no longer apply.

In this paper, we proposed a time-section fusion pattern classification based day-ahead solar irradiance ensemble forecasting model using mutual iterative optimization. The historical solar irradiance data is first decomposed into several sub-sequences using wavelet decomposition (WD) at different levels. Then, multiple WD based solar irradiance forecasting models are applied to achieve multiple solar irradiance forecasting results. Then, the fusion pattern classification model is used to recognize the fusion pattern of data at each time section. Finally, the forecasting result can be achieved by fuse the multiple forecasting outputs into one individual value using the fusion model corresponding to the fusion pattern.

To ensure the model performance, a mutual iterative optimization framework is proposed to optimize the fusion pattern classification model and corresponding data fusion models.

The main contributions of this paper include:

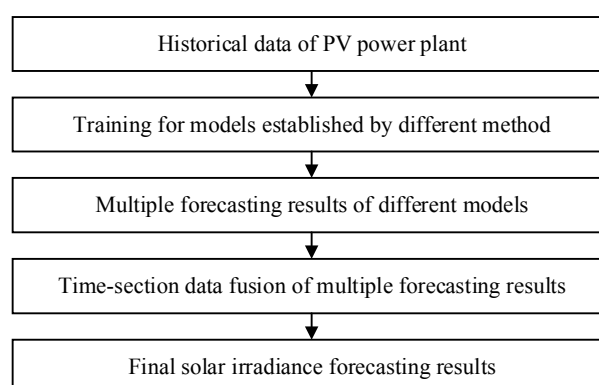
- (1) We propose a time-section fusion pattern classification based day-ahead solar irradiance ensemble forecasting model.
- (2) We propose a mutual iterative optimization framework to optimize the pattern classification and data fusion models.
- (3) We test the performance of the proposed ensemble forecasting model and improve the accuracy of solar irradiance day-ahead forecasting approach.

The rest of this paper is organized as follows. Section 2 introduces the ensemble model for day-ahead solar irradiance forecasting. Section 3 introduces the proposed mutual iterative optimization framework for pattern classification and data fusion models. Section 4 is the simulation part. Finally, conclusions were drawn in Section 5.

## 2. Ensemble Model for Day-Ahead Solar Irradiance Forecasting

### 2.1. Ensemble Forecasting Framework

In consideration of the diversity of solar irradiance variation under different weather conditions, using a single model to forecast irradiance data is inadvisable. Ensemble forecasting with multiple models established by different method or theory has been proved as an effective means. Therefore, a framework of ensemble model for day-ahead solar irradiance forecasting is applied in this paper, as shown in Figure 1.



**Figure 1.** The framework of day-ahead solar irradiance ensemble forecasting.

The proposed ensemble forecasting model contains two processes. In the first process, multiple forecasting models are established according to different forecasting methods or theories. After all the forecasting models are trained using historical data of PV power plant, we can achieve multiple solar irradiance forecasting results at each time section in the forecasting day. Then, in the second process, to obtain the final forecasting results, a fusion model is needed to fuse all of the multiple forecasting outputs into a single irradiance value at each time section.

## 2.2. Wavelet Decomposition Based Irradiance Forecasting

Wavelet theory provides an efficient tool for complex data sequence analysis due to its advantages in multi-scale information processing [23,24]. By discrete wavelet transform (DWT), the original data series can be decomposed into two parts called approximate component and detailed component. The approximate contains the low-frequency information of original series, while the detailed component is focused on high-frequency information. This process is known as wavelet decomposition (WT), and the two subseries obtained from the original series can also be further decomposed by WD process. Therefore, the fluctuant and random part of a data series, which is usually considered as the high-frequency noise, can be extracted and filtered using WD process. Based on the above theory, an efficient way to help improving the forecasting accuracy under non-stationary weather statuses with high volatility data is to decompose the original data series into several stable parts and fluctuant parts. These decomposed subseries have better behaviors (e.g., more stable variances and fewer outliers) in terms of regularity than the original data series, and thus can be forecasted more accurately using multiple well-directed models [25–30].

Given a certain mother wavelet  $\psi(t)$  and its corresponding scaling function  $\varphi(t)$ , a sequence of wavelet  $\psi_{j,k}(t)$  and binary scale-functions  $\varphi_{j,k}(t)$  can be developed as follows:

$$\psi_{j,k}(t) = 2^{\frac{j}{2}} \psi(2^j t - k) \quad (1)$$

$$\varphi_{j,k}(t) = 2^{\frac{j}{2}} \varphi(2^j t - k) \quad (2)$$

where  $t$  is the time index,  $j$  and  $k$  denote the scaling and translation variables, respectively.

Then the original data sequence  $s(t)$  can be presented as:

$$s(t) = \sum_{k=1}^n c_{j,k} \varphi_{j,k}(t) + \sum_{j=1}^J \sum_{k=1}^n d_{j,k} \psi_{j,k}(t) \quad (3)$$

where  $c_{j,k}$  represents the approximation coefficient at scale  $j$  and location  $k$ ,  $d_{j,k}$  represents the detail coefficient at scale  $j$  and location  $k$ ,  $n$  is the size of the data sequence, and  $J$  is the decomposition level. Therefore, according to the fast DWT algorithm, which was developed by Mallat [31], the approximate component and detailed component of a certain WD level can be obtained through multiple low-pass filters (LPF) and high-pass filters (HPF) [30,32,33].

As shown in Figure 2, the original data sequence  $S$  can be first decomposed into two part: approximate component  $A1$  and detailed component  $D1$  at WD level 1. Then, the approximate component can also be decomposed into a secondary approximate component  $A2$  and detailed component  $D2$  at WD level 2, and then continues to  $A3$  and  $D3$  at WD level 3, etc.

Therefore, for a certain WD level  $k$ , the total subsequences obtained from original sequence  $S$  through the WD process are approximate subsequence  $A_k$  and detailed subsequences  $D1$  to  $D_k$ . Based on the above theory, if the level of WD is determined as  $k$ , then the subsequences of approximations  $A_k$  and details from  $D1$  to  $D_k$  can be calculated by DWT. All of these subsequences can be forecasted individually using some kind of time sequence forecasting theory, such as autoregressive (AR) models, Artificial Neural Network (ANN), and support vector regression (SVR). Then final forecasting results of original data sequence are calculated by inverse discrete wavelet transform

(IDWT) of the forecasting results of  $A_k$  and  $D_1$  to  $D_k$ . The process of this  $k$ -level WD based solar irradiance forecasting method is shown in Figure 3.

In this paper, the multiple forecasting models for ensemble model can be achieved using the WD based solar irradiance forecasting method with different values of WD level  $k$ .

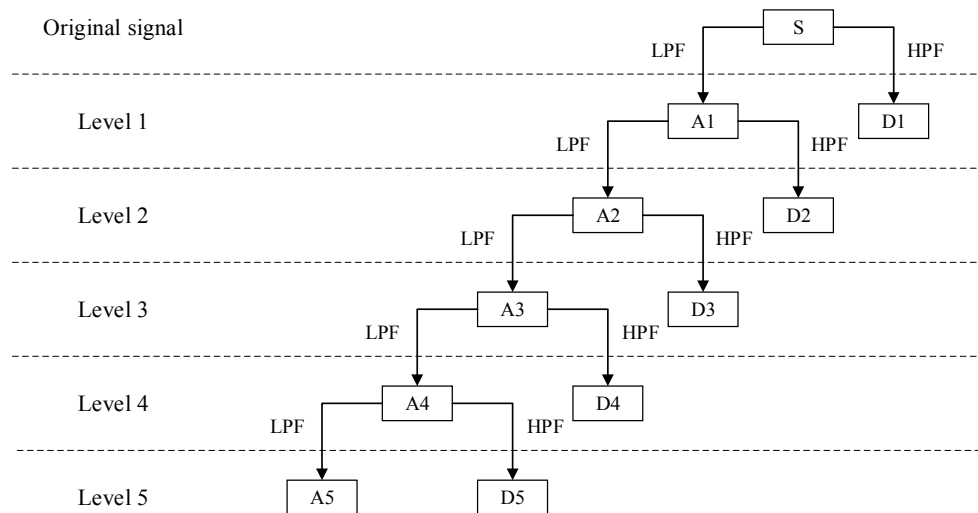


Figure 2. Wavelet decomposition tree.

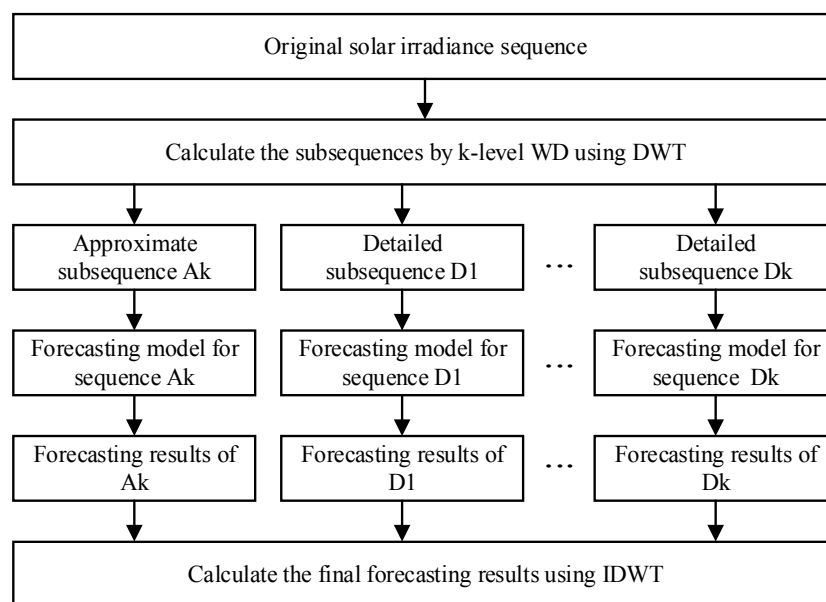


Figure 3. The  $k$ -level wavelet decomposition (WD) based solar irradiance forecasting method.

### 2.3. Time-Section Data Fusion and Fusion Pattern Classification

Based on the  $k$ -level WD based time sequence forecasting method described above in Section 2.2, six individual irradiance day-ahead forecasting models are built and trained using original irradiance sequence and decomposed irradiance sub-sequences at WD level 1 to 5.

In the case of the structural and parameters of the above six models are determined, the following data can be obtained in simulation:

- (1) Solar irradiance forecasting results using original data:  $IRR_{p0}$ ;
- (2) Solar irradiance forecasting results using decomposed data:  $IRR_{p1}$ ,  $IRR_{p2}$ ,  $IRR_{p3}$ ,  $IRR_{p4}$ , and  $IRR_{p5}$ ;

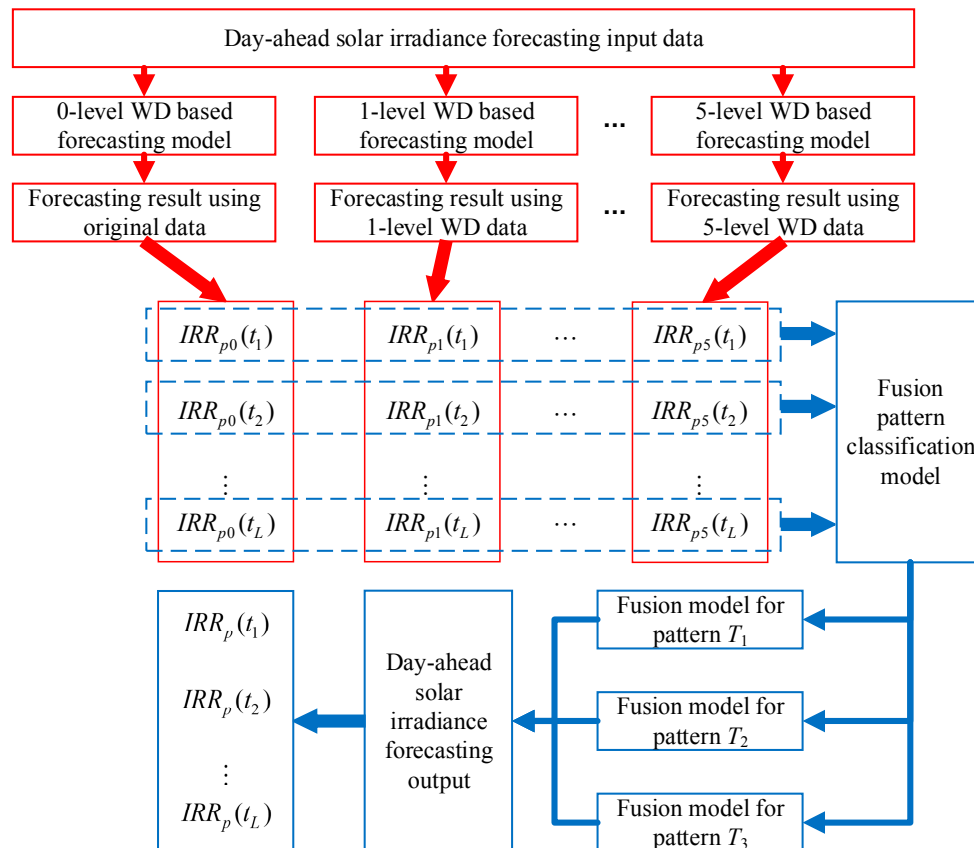
(3) Actual solar irradiance data:  $IRR_o$ ;

The three kinds of data can be exhibited in the matrix:

$$\begin{Bmatrix} IRR_{p0}(t_1) & IRR_{p1}(t_1) & IRR_{p2}(t_1) & IRR_{p3}(t_1) & IRR_{p4}(t_1) & IRR_{p5}(t_1) & IRR_o(t_1) \\ IRR_{p0}(t_2) & IRR_{p1}(t_2) & IRR_{p2}(t_2) & IRR_{p3}(t_2) & IRR_{p4}(t_2) & IRR_{p5}(t_2) & IRR_o(t_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ IRR_{p0}(t_L) & IRR_{p1}(t_L) & IRR_{p2}(t_L) & IRR_{p3}(t_L) & IRR_{p4}(t_L) & IRR_{p5}(t_L) & IRR_o(t_L) \end{Bmatrix}$$

Then for each time section  $t$ , a fusion model is needed to fuse irradiance data  $IRR_{p0}(t)$ ,  $IRR_{p1}(t)$ ,  $IRR_{p2}(t)$ ,  $IRR_{p3}(t)$ ,  $IRR_{p4}(t)$ , and  $IRR_{p5}(t)$  to  $IRR_o(t)$ . However, as the characteristic of solar irradiance changes with the time, the performance of each forecasting model will change correspondingly as well. The mapping relationship between different forecasting results and actual irradiance is also unstable, which means that there will be multiple fusion patterns for forecasting results at different time sections. Therefore, to classify all of the data into different classes according to their fusion patterns and build the corresponding fusion model can achieve a better result than using only one single fusion model.

According to the above analysis, a time-section fusion pattern classification based day-ahead solar irradiance ensemble forecasting model is proposed in this paper. The overall framework of the proposed model is shown in Figure 4. In this exhibit framework, six forecasting models are applied according to different WD level, and the number of fusion pattern is set as three.



**Figure 4.** The framework of the proposed time-section fusion pattern classification based day-ahead solar irradiance ensemble forecasting model.

According to the forecasting process of proposed model, in the case of determined structure and parameters of six forecasting models, the performance of fusion pattern classification model, targeted

fusion models, and the coordination between the two kinds of models can greatly influence the final solar irradiance forecasting accuracy, which will be further discussed and analyzed in the next section.

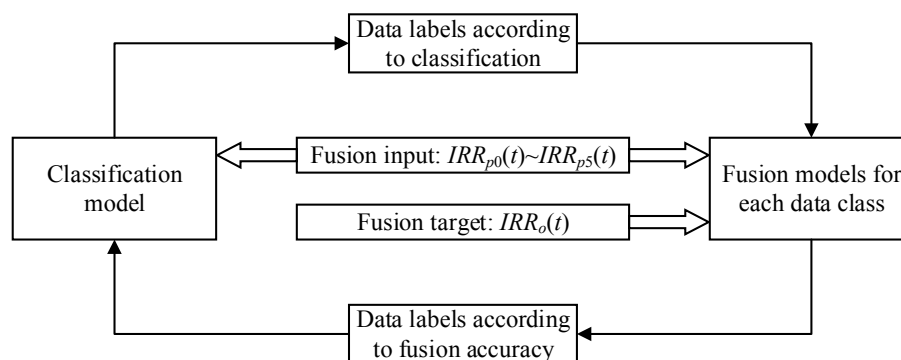
### 3. Mutual Iterative Optimization for Classification and Fusion Models

#### 3.1. Methodology

An optimal fusion pattern classification model and fusion model corresponding to each fusion pattern are very necessary for the accuracy of our proposed time-section fusion pattern classification based day-ahead solar irradiance ensemble forecasting model. Then, to distinguish different fusion patterns and build the corresponding fusion model automatically, a mutual iterative optimization framework is proposed in this paper.

As the performances of the two kinds of models are inter-related, the final output of forecasting model not only depends on the performance of the two kinds of models, but also the cooperation efficiency between them. Therefore, the fundamental of the proposed mutual iterative optimization framework is based on the idea that using the results of pattern classification model to guide the training of fusion models, and then go back update the classification model according to the fusion accuracy.

This constitutes a calculation cycle, as shown in Figure 5, in which the classification and fusion models are unstable as they will be modified by each other and keep updating constantly. Only when the two kinds of models are perfectly coordinated, thus the data labels according to classification and according to fusion accuracy are exactly the same, can the cycle reach convergence. Then, the needed pattern classification and corresponding data fusion models in our day-ahead solar irradiance ensemble forecasting model are obtained as well.



**Figure 5.** The fundamental of the proposed mutual optimization framework for pattern classification and data fusion models.

#### 3.2. Algorithm Procedures

Based on the above analysis, the detail procedures of the mutual iterative optimization framework are designed as follows:

- (1) Set the counting unit  $k = 1$ . Cluster all of the forecasting results of multiple models at each time section into  $n$  classes using k-means algorithm.
- (2) For data in each pattern class  $C_i$  ( $i = 1, 2, \dots, n$ ), build and train the fusion model  $F_i^k$  using forecasting results (fusion model input) and actual data (fusion model output).
- (3) Fuse all the forecasting results using each fusion model  $F_i^k$ , reclassify the data into different pattern classes according to fusion accuracy. Thus, if the fusion result of model  $F_i^k$  is closest to the actual data at time section  $t$ , then reset the pattern label  $L(t)$  to  $C_i$ .
- (4) According to the forecasting results and their new pattern labels, build and train a data classification model  $F_c^k$ .
- (5) Check if the iteration termination condition is satisfied.



- (6) If no, using model  $F_c^k$  to reclassify the data, then goes back to procedure (2). If yes, output the classification model and fusion models as optimized models.

As Figure 6 shows, during the optimization process, the classification model and fusion models are modified mutually in each cycle. Then, the structure and parameters of classification and fusion models are recorded and updated if the performance of final day-ahead solar irradiance forecasting model is increased after a new cycle. The iteration termination condition is set as: no new classification and fusion models are recorded in 200 consecutive cycles or the total number of cycles reaches 2000.

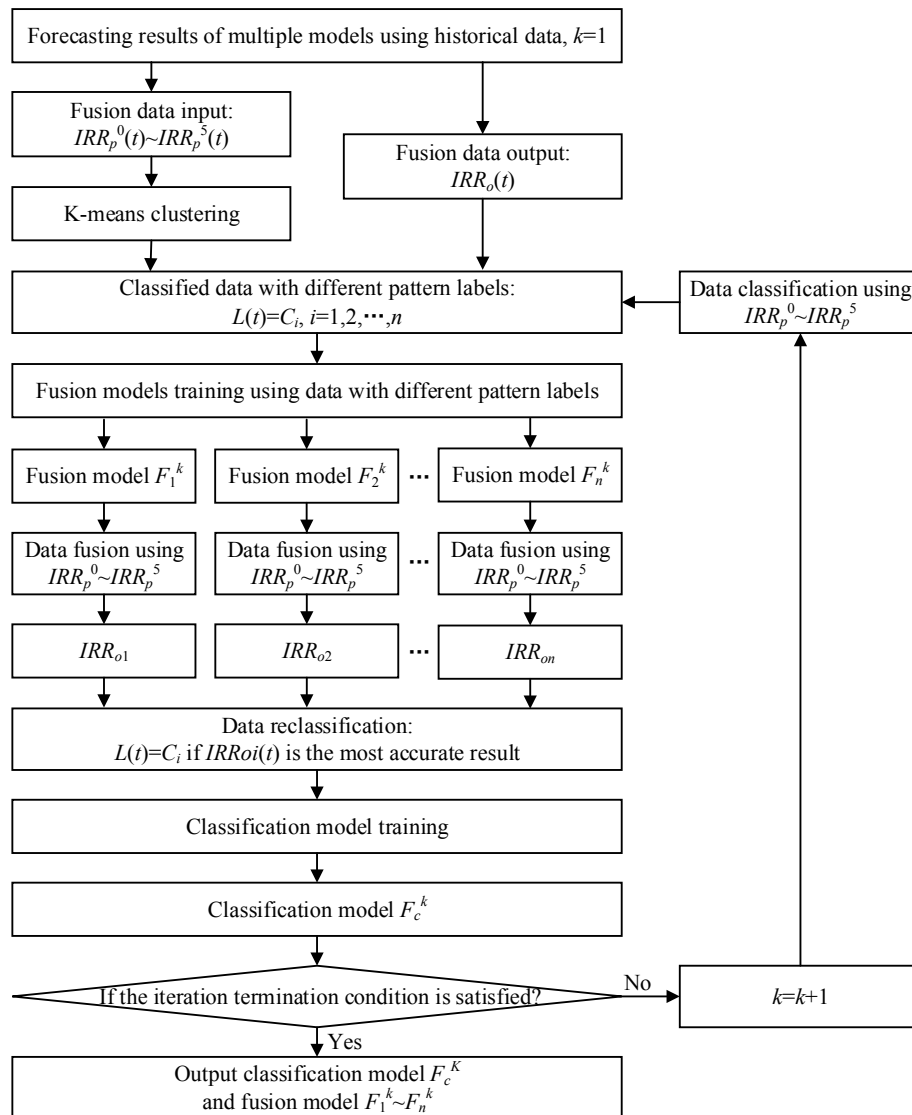


Figure 6. The framework of mutual optimization for pattern classification and data fusion models.

## 4. Simulation and Discussion

### 4.1. Data

The solar irradiance data for our simulation is downloaded on the National Oceanic & Atmospheric Administration (NOAA) Earth System Research Laboratory website, measured in the Surface Radiation (SURFRAD) station at Desert Rock and Sioux Falls from 2014 to 2016 [34]. There are totally 967 days with available data at Desert Rock and 1021 days with available data at Sioux Falls as shown in Figures 7 and 8. The time resolution of original solar irradiance data is 1 min.



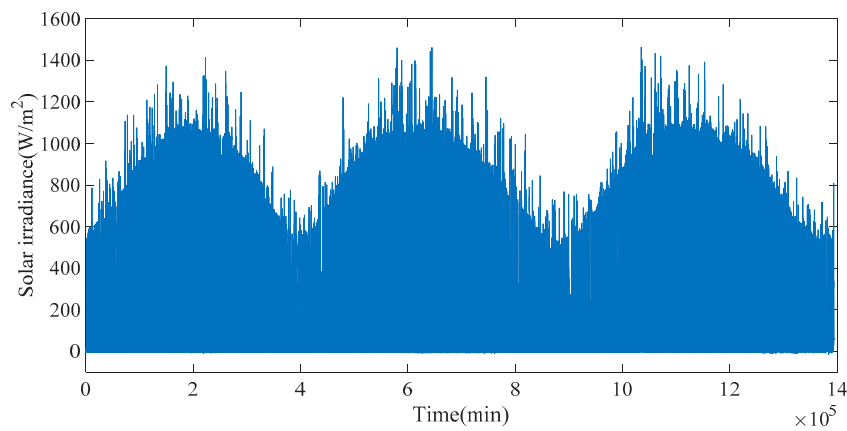
The simulation platform is the MATLAB software (version R2015b), which executed on a laptop PC with a Core(TM) CPU running at 2.3 GHz and a memory capacity of 8 GB under Windows 10 Operating System. To evaluate the forecasting accuracy, three error indexes of Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Correlation Coefficient (COR) are used to evaluate the accuracy.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (4)$$

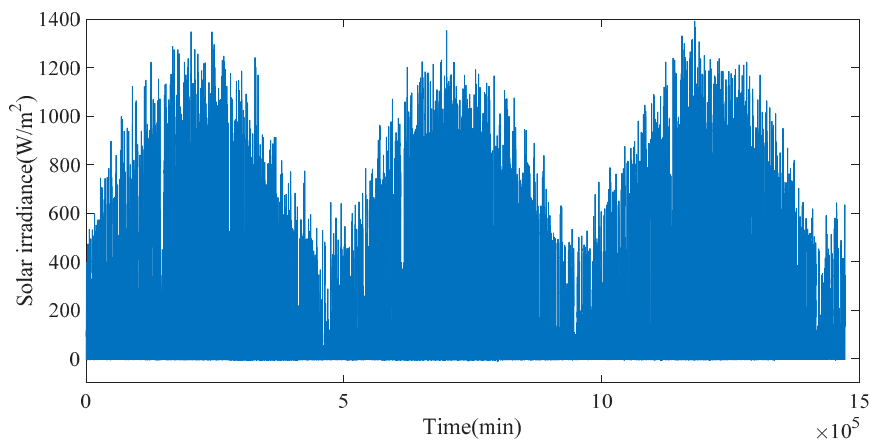
$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (5)$$

$$\text{COR} = \frac{\text{Cov}(y, \hat{y})}{\sqrt{V(y)}\sqrt{V(\hat{y})}} \quad (6)$$

where  $y$  is the actual value,  $\hat{y}$  is the forecasted value,  $n$  is the number of forecasting results.



**Figure 7.** Solar irradiance data of Desert Rock station from 2014 to 2016.



**Figure 8.** Solar irradiance data of Sioux Falls station from 2014 to 2016.

#### 4.2. Simulation Design

When considering the limitation of computing ability and the need for day-ahead forecasting in simulation, the time interval of irradiance sequence is changed from 1 min to 1 h, i.e., the new hourly data is obtained by the average value of every 60 data series within one hour in the original sequence.

Six models are applied to forecast the solar irradiance respectively a standard artificial neural network (ANN) fitting model for irradiance forecasting, and five WD based irradiance forecasting models with WD level from 1 to 5 (described in Section 2.2). The forecast of decomposed sequences in

WD based model is also realized by the ANN fitting model. The solar irradiance data in the years of 2014 and 2015 are used as training set and the data in the year of 2016 is used as a testing set for the six forecasting models.

Then, for the pattern classification and data fusion models, we set the number of fusion patterns as 2 and select 1000 continuous time sections start at different times to simulate the model, with the former 800 data for training and optimization and the last 200 data for testing. The pattern classification and data fusion are also realized according to ANN.

The accuracy of final forecasting results of solar irradiance will be compared with the output of the first six forecasting models (i.e., standard ANN fitting model and five WD based models), to evaluate the effectiveness of proposed time-section fusion pattern classification based ensemble forecasting model framework.

The reason for choosing the same ANN to realize the fitting, classification, and fusion, on one hand, is because the main contribution of this paper is the proposed forecasting framework utilizing the data fusion pattern classification, using models that are based on the same ANN algorithm can avoid the results deviations caused by the differences in learning abilities of different models. On the other hand, is because ANN is a mature machine learning algorithm and can adapt to different application requirements.

#### 4.3. Simulation Results and Comparison

The solar irradiance data in the last 200 days of 2016 at Desert Rock station is forecasted using eight models: standard ANN model, WD based forecasting models from WD level 1 to 5, an ensemble model using a single fusion model to fuse the 6 models outputs into one result, and the proposed fusion pattern classification based ensemble forecasting model. The deviation between actual solar irradiance and forecasting results of different models in the last month of 2016 from 6:00 a.m. to 6:00 p.m. are shown in Figure 9.

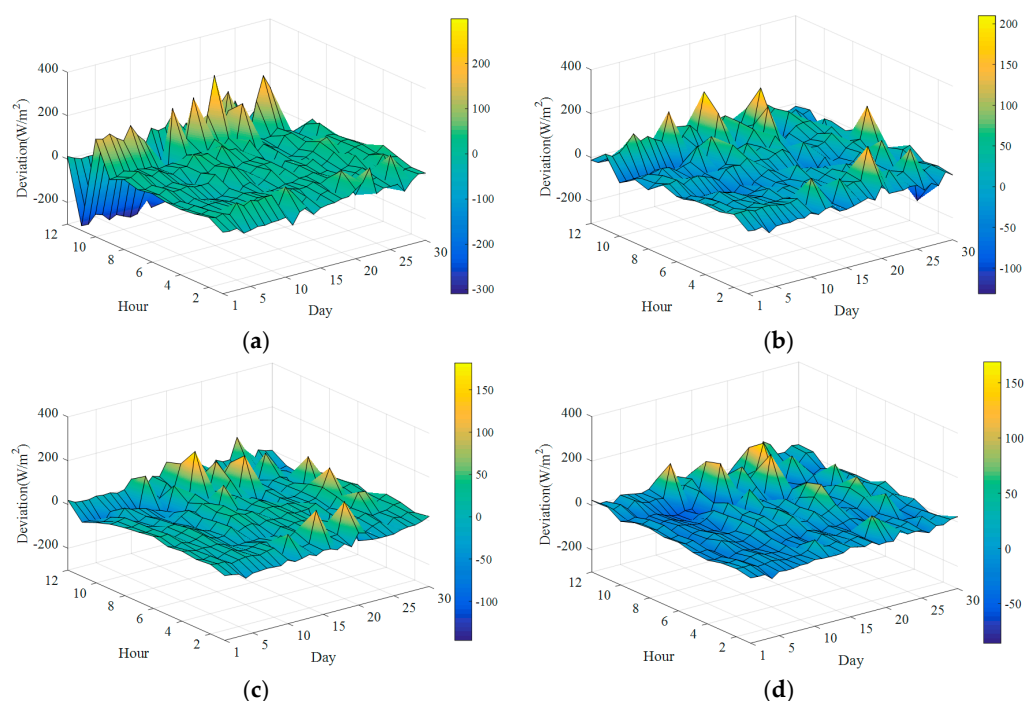
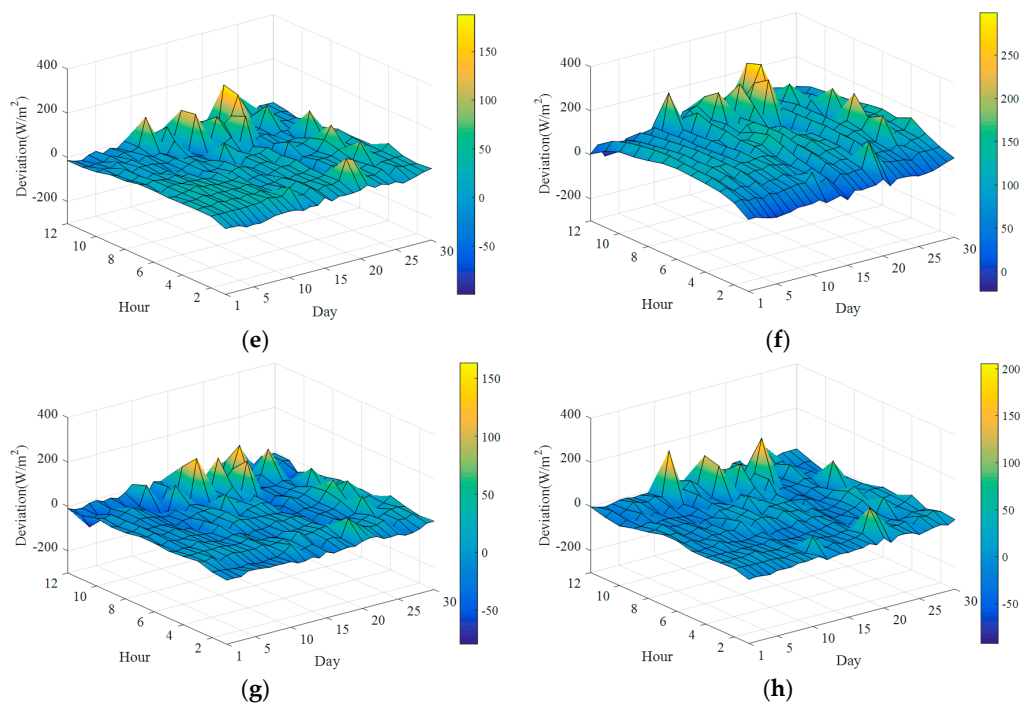
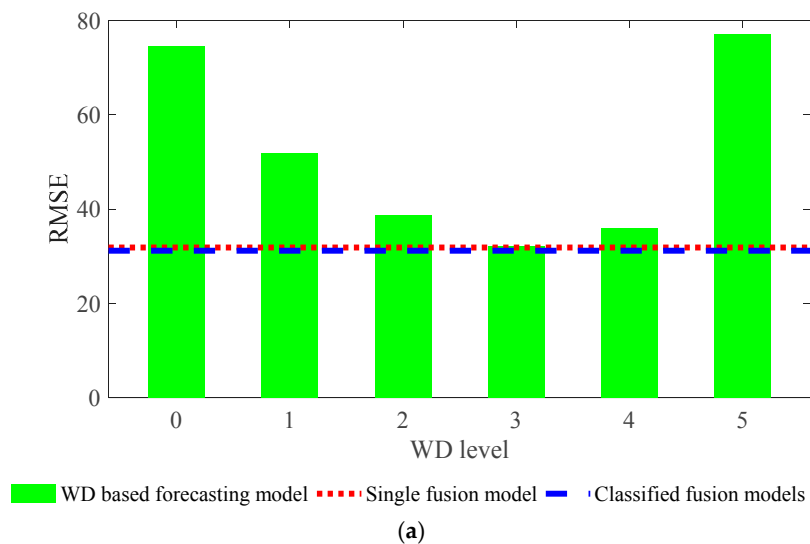


Figure 9. Cont.

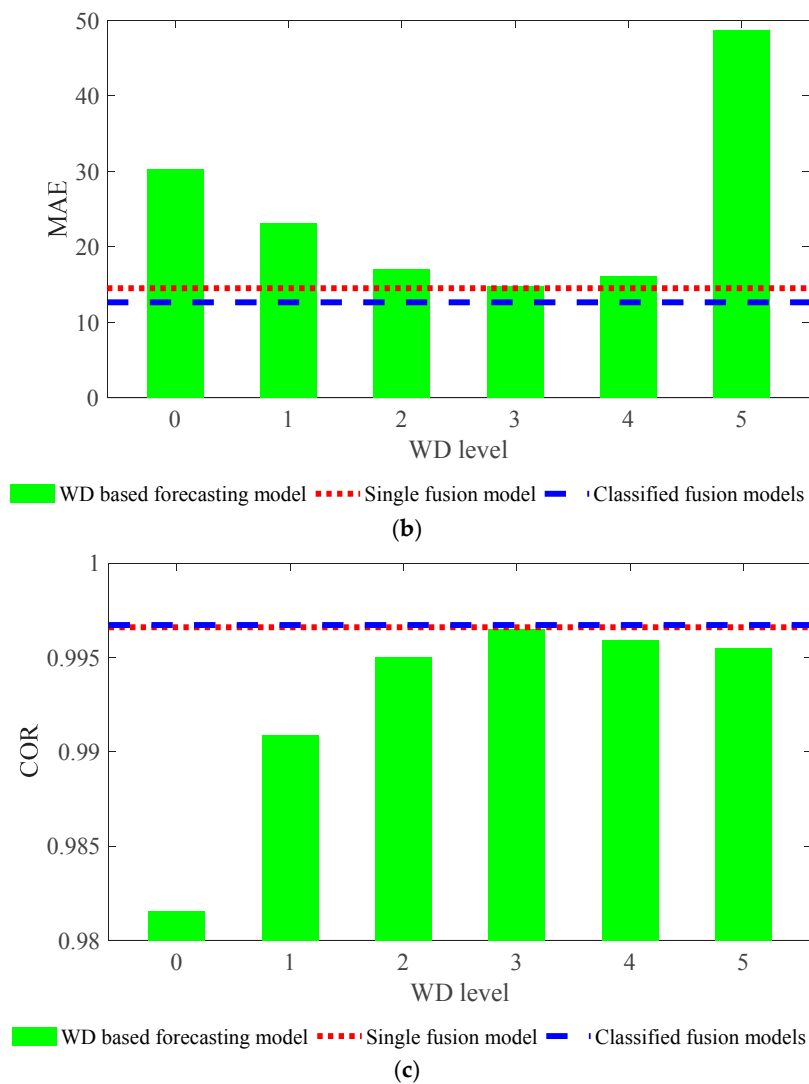


**Figure 9.** The deviation between actual and forecasted solar irradiance of different models. (a) WD level 0; (b) WD level; (c) WD level 2 (d); WD level 3; (e) WD level 4; (f) WD level 5; (g) Single fusion model; and, (h) Classified fusion models.

Then the overall accuracy of the 8 models is calculated and shown in Figure 10. As the WD based models also utilize ANN to realize data forecasting, the standard ANN forecasting model can be also considered as a WD based model with WD level 0.



**Figure 10.** Cont.

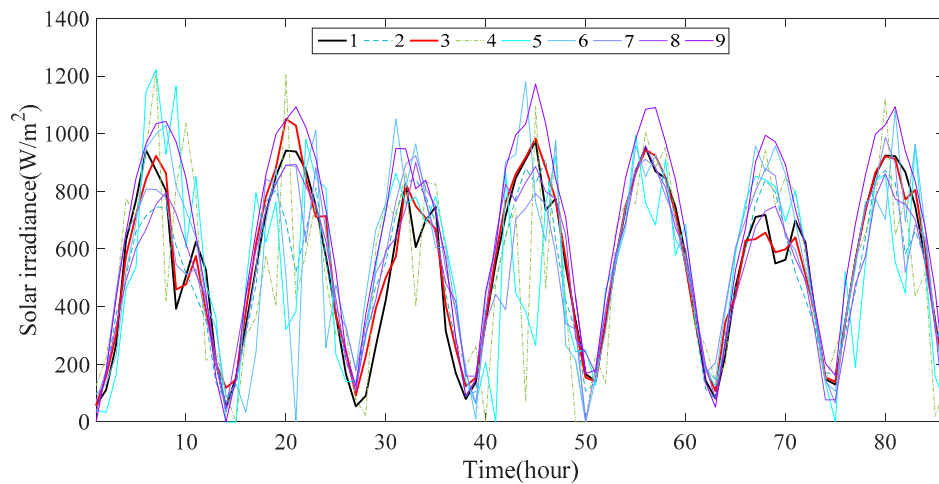


**Figure 10.** The accuracy of different forecasting models at Desert Rock station. (a) The Root Mean Squared Error (RMSE) of forecasting results; (b) The Mean Absolute Error (MAE) of forecasting results; and, (c) The Correlation Coefficient (COR) of forecasting results.

It can be seen that, for the six WD based forecasting models, the overall accuracy of the WD level 3 model is the highest. Then, by fusing all of the forecasting results into one irradiance value at each time section, the ensemble model using a single fusion model can achieve a better accuracy than any individual WD based model.

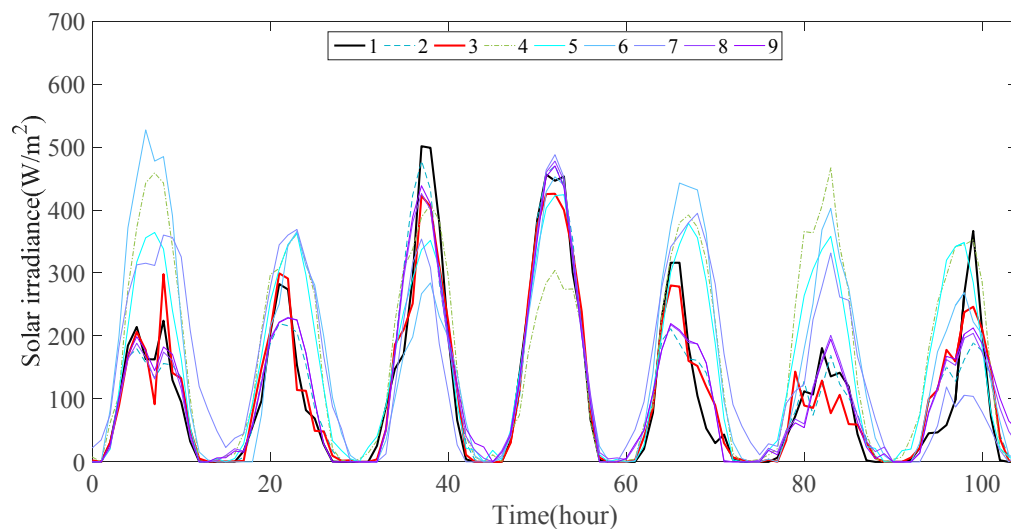
Moreover, on this basis, the proposed time-section fusion pattern classification based day-ahead solar irradiance ensemble forecasting model achieved a further more accurate irradiance forecasting result, which proved the effectiveness of the proposed mutual iterative optimization algorithm.

The detail forecasting results of the eight models in continues seven days are shown in Figure 11, solar irradiance data at the time before sunrise and after sunset is neglected in the figure.



**Figure 11.** Actual and forecasted solar irradiance at Desert Rock station. Figure legend: 1-Actual data, 2-Forecast results using single fusion model, 3-Forecast results using classified fusion models, 4-Forecast results of 0-level WD model (standard ANN model), 5-Forecast results of 1-level WD model, 6-Forecast results of 2-level WD model, 7-Forecast results of 3-level WD model, 8-Forecast results of 4-level WD model, 9-Forecast results of 5-level WD model.

According to the actual solar irradiance curve in Figure 11, there are multiple weather conditions, including sunny and cloudy days among the 10 days, and the output of the proposed fusion pattern classification based ensemble forecasting model can track the actual irradiance curve under these different weather conditions. We also simulate and test the models using Sioux Falls station's data. Continues 10 days forecasting results under different weather conditions are shown in Figure 12.



**Figure 12.** Actual and forecasted solar irradiance at Sioux Falls station. Figure legend: 1-Actual data, 2-Forecast results using single fusion model, 3-Forecast results using classified fusion models, 4-Forecast results of 0-level WD model (standard ANN model), 5-Forecast results of 1-level WD model, 6-Forecast results of 2-level WD model, 7-Forecast results of 3-level WD model, 8-Forecast results of 4-level WD model, 9-Forecast results of 5-level WD model.

It can be seen that the meteorological conditions of Sioux Falls station are much more unstable than Desert Rock station. Namely, its daily solar irradiance curves show more diversity, according to the measuring database in NOAA Earth System Research Laboratory. However, the proposed

fusion pattern classification based ensemble model still shows higher accuracy than the other seven forecasting models even under this condition.

To further test the model performance in different weather conditions, the forecasting results of Sioux Falls station are separated into two parts based on cloudiness (clear sky or cloudy), the evaluated accuracy of the eight models are shown in Table 1. It can be seen that in cloudy days, the proposed model can achieve more improvement in forecasting results than in clear sky days.

**Table 1.** The performance of models under different weather conditions.

| Forecasting Models         |                         | RMSE           |             |
|----------------------------|-------------------------|----------------|-------------|
|                            |                         | Clear Sky Days | Cloudy Days |
| WD based forecasting model | WD level 0              | 79.57          | 112.40      |
|                            | WD level 1              | 101.76         | 109.44      |
|                            | WD level 2              | 55.34          | 120.43      |
|                            | WD level 3              | 23.16          | 100.19      |
|                            | WD level 4              | 23.06          | 46.02       |
|                            | WD level 5              | 17.32          | 46.23       |
| Ensemble forecasting model | Single fusion model     | 19.55          | 43.78       |
|                            | Classified fusion model | 16.76          | 36.83       |

In the end, to evaluate the performance of the proposed method when compared with other common forecasting methods, the persistence model and autoregressive integrated moving average (ARIMA) model are also applied to forecast the irradiance. For persistence model, we use the solar irradiance data in the day as the day-ahead forecasting results of the next day directly. For the ARIMA model, an ARIMA (3,1,2) model is built and trained to forecast the next day's irradiance data according to the previous three days' data. The values of different models' accuracy at Desert Rock station and Sioux Falls station are shown in Tables 2 and 3. The comparison shows that the proposed time-section fusion pattern classification based day-ahead solar irradiance ensemble forecasting model is much more accurate than traditional ANN model (the WD level 0 model in this paper), persistence model, and ARIMA model.

**Table 2.** The accuracy of different forecasting models at Desert Rock station.

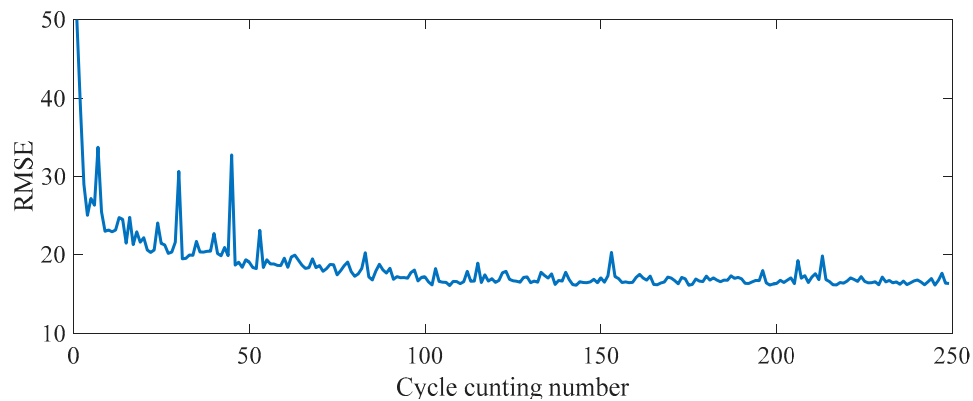
| Model | WD Based Forecasting Model |            |            |            |            |            | Ensemble Model      |                          | Persistence Model | ARIMA Model |
|-------|----------------------------|------------|------------|------------|------------|------------|---------------------|--------------------------|-------------------|-------------|
|       | WD Level 0                 | WD Level 1 | WD Level 2 | WD Level 3 | WD Level 4 | WD Level 5 | Single Fusion Model | Classified Fusion Models |                   |             |
| RMSE  | 74.66                      | 51.86      | 38.66      | 32.10      | 36.07      | 76.99      | 31.81               | 31.15                    | 49.83             | 112.19      |
| MAE   | 30.33                      | 23.09      | 16.96      | 14.81      | 16.04      | 48.70      | 14.50               | 12.62                    | 19.41             | 62.10       |
| COR   | 0.9815                     | 0.9908     | 0.9950     | 0.9965     | 0.9959     | 0.9955     | 0.9965              | 0.9967                   | 0.9916            | 0.9589      |

**Table 3.** The accuracy of different forecasting models at Sioux Falls station.

| Model | WD Based Forecasting Model |            |            |            |            |            | Ensemble Model      |                          | Persistence Model | ARIMA Model |
|-------|----------------------------|------------|------------|------------|------------|------------|---------------------|--------------------------|-------------------|-------------|
|       | WD Level 0                 | WD Level 1 | WD Level 2 | WD Level 3 | WD Level 4 | WD Level 5 | Single Fusion Model | Classified Fusion Models |                   |             |
| RMSE  | 102.45                     | 106.22     | 103.14     | 82.26      | 38.61      | 37.98      | 31.81               | 31.15                    | 125.01            | 143.98      |
| MAE   | 68.56                      | 69.63      | 63.73      | 52.41      | 24.72      | 23.80      | 14.50               | 12.62                    | 70.72             | 84.21       |
| COR   | 0.8322                     | 0.8113     | 0.8500     | 0.9075     | 0.9777     | 0.9782     | 0.9965              | 0.9967                   | 0.7059            | 0.5517      |

#### 4.4. Simulation Discussion

In the performed simulation of Desert Rock station, the counting unit  $k$  reached 249, and the test performance (here use the RMSE as performance index) of proposed day-ahead solar irradiance ensemble forecasting model after each cycle is recorded, as shown in Figure 13. The accuracy of the ensemble forecasting model increases rapidly in the first 100 cycles. When the cycle counting number is over 100, the value of RMSE began to stabilize.



**Figure 13.** The validation performance of the proposed model after each cycle.

It is noteworthy that although the trend of model performance keeps increasing, the RMSE value may rebound in several cycles. This may be because that the classification and fusion models are realized by ANN algorithm and the random factors in the initialization of ANN can cause some incongruity between the two kinds of models. During the completely cyclic iterative optimization process, the minimum of validation RMSE value can reach to 16.1016, which is much lower than the RMSE value of forecasting results obtained by the proposed model.

On the one hand, although the two are positively related, in practical application the validation RMSE is calculated using available historical data to check the model's ability to learn and match data, which is different from the RMSE used to evaluate the model forecasting performance. On the other hand, the surrounding environment of photovoltaic power stations is changing over time, the model trained and optimized using historical data may not suit the new characteristics of irradiance. Therefore, it is also necessary to update the model regularly.

## 5. Conclusions

For the proposed time-section fusion pattern classification based day-ahead solar irradiance ensemble forecasting model, once the forecasting part models are determined according to historical training data, the fusion pattern classification model and fusion models corresponding to each fusion pattern will be the most important factors of the forecasting approach. Not only can it affect the performance of each individual classification and fusion model, but also the compatibility of these two kinds of models can affect the final forecasting accuracy. Therefore, to improve the performance of the day-ahead solar irradiance ensemble forecasting model, a mutual iterative optimization framework for pattern classification and data fusion models is proposed in this paper as well. Simulations showed that the pattern classification model and fusion models interact with each other through the fusion pattern labels in the optimization process and the model performance increased after most optimization cycles. The two kinds of models will continue to modify each other during the calculation cycle until reaching the absolute coordination. A comparison of forecasting accuracy was made between each single level WD base forecasting model, the ensemble model with single fusion model, and the ensemble model with pattern classification and data fusion models optimized by the proposed mutual iterative optimization framework. Finally, the effectiveness of the proposed time-section fusion pattern classification based day-ahead solar irradiance ensemble forecasting model using mutual iterative optimization was proved by the comparative results.

**Acknowledgments:** This work was supported partially by the National Natural Science Foundation of China (Grant No. 51577067), the National Key Research and Development Program of China (Grant No. 2017YFF0208106), the Beijing Natural Science Foundation of China (Grant No. 3162033), the Beijing Science and Technology Program of China (Grant No. Z161100002616039), the Hebei Natural Science Foundation of China (Grant No. E2015502060), the State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources (Grant No. LAPS16007, LAPS16015), the Science and Technology Project of State Grid Corporation



of China (SGCC), the Open Fund of State Key Laboratory of Operation and Control of Renewable Energy & Storage Systems (China Electric Power Research Institute) (No. 5242001600FB) and the China Scholarship Council. M. Shafie-khah and J.P.S. Catalão acknowledge the support by FEDER funds through COMPETE 2020 and by Portuguese funds through FCT, under Projects SAICT-PAC/0004/2015—POCI-01-0145-FEDER-016434, POCI-01-0145-FEDER-006961, UID/EEA/50014/2013, UID/CEC/50021/2013, and UID/EMS/00151/2013, and also funding from the EU 7th Framework Programme FP7/2007-2013 under GA No. 309048.

**Author Contributions:** All authors have worked on this manuscript together and all authors have read and approved the final manuscript.

**Conflicts of Interest:** The authors declare that the grant, scholarship, and/or funding mentioned in the Acknowledgments section do not lead to any conflict of interest. Additionally, the authors declare that there is no conflict of interest regarding the publication of this manuscript.

## References

1. Chu, Y.; Urquhart, B.; Gohari, S.M.I.; Pedro, H.T.C.; Kleissl, J.; Coimbra, C.F.M. Short-term reforecasting of power output from a 48 MWe solar PV plant. *Sol. Energy* **2015**, *112*, 68–77. [CrossRef]
2. Seyboth, K.; Sverrisson, F.; Appavou, F.; Brown, A.; Epp, B.; Leidreiter, A.; Lins, C.; Musolino, E.; Murdock, H.E.; Petrichenko, K.; et al. *Renewables 2016 Global Status Report*; REN21: Paris, France, 2016.
3. 2015 PV-Related Statistics. Available online: [http://www.nea.gov.cn/2016-02/05/c\\_135076636.htm](http://www.nea.gov.cn/2016-02/05/c_135076636.htm) (accessed on 22 October 2017).
4. Wang, F.; Xu, H.; Xu, T.; Li, K.; Shafie-Khah, M.; Catalão, J.P.S. The values of market-based demand response on improving power system reliability under extreme circumstances. *Appl. Energy* **2017**, *193*, 220–231. [CrossRef]
5. Tuohy, A.; Zack, J.; Haupt, S.E.; Sharp, J.; Ahlstrom, M.; Dise, S.; Gritmit, E.; Mohrlen, C.; Lange, M.; Casado, M.G.; et al. Solar Forecasting: Methods, Challenges, and Performance. *IEEE Power Energy Mag.* **2015**, *13*, 50–59. [CrossRef]
6. Chen, Q.; Wang, F.; Hodge, B.-M.; Zhang, J.; Li, Z.; Shafie-Khah, M.; Catalao, J.P.S. Dynamic Price Vector Formation Model Based Automatic Demand Response Strategy for PV-assisted EV Charging Station. *IEEE Trans. Smart Grid* **2017**, *8*, 2903–2915. [CrossRef]
7. Wang, F.; Zhou, L.; Ren, H.; Liu, X.; Talari, S.; Shafie-Khah, M.; Catalao, J.P.S. Multi-objective Optimization Model of Source-Load-Storage Synergetic Dispatch for Building Energy System Based on TOU Price Demand Response. *IEEE Trans. Ind. Appl.* **2017**. [CrossRef]
8. 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]
9. Antonanzas, J.; Osorio, N.; Escobar, R.; Urraca, R.; Martinez-de-Pison, F.J.; Antonanzas-Torres, F. Review of photovoltaic power forecasting. *Sol. Energy* **2016**, *136*, 78–111. [CrossRef]
10. 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]
11. Sun, Y.; Wang, F.; Wang, B.; Chen, Q.; Engerer, N.A.; Mi, Z. Correlation Feature Selection and Mutual Information Theory Based Quantitative Research on Meteorological Impact Factors of Module Temperature for Solar Photovoltaic Systems. *Energies* **2016**, *10*, 7. [CrossRef]
12. Monteiro, C.; Santos, T.; Fernandez-Jimenez, L.A.; Ramirez-Rosado, I.J.; Terreros-Olarte, M.S. Short-term power forecasting model for photovoltaic plants based on historical similarity. *Energies* **2013**, *6*, 2624–2643. [CrossRef]
13. Wang, F.; Mi, Z.; Su, S.; Zhao, H. Short-term solar irradiance forecasting model based on artificial neural network using statistical feature parameters. *Energies* **2012**, *5*, 1355–1370. [CrossRef]
14. Marquez, R.; Coimbra, C.F.M. Intra-hour DNI forecasting based on cloud tracking image analysis. *Sol. Energy* **2013**, *91*, 327–336. [CrossRef]
15. Chow, C.W.; Belongie, S.; Kleissl, J. Cloud motion and stability estimation for intra-hour solar forecasting. *Sol. Energy* **2015**, *115*, 645–655. [CrossRef]
16. Wang, F.; Zhen, Z.; Liu, C.; Mi, Z.; Hodge, B.-M.; Shafie-Khah, M.; Catalão, J.P.S. Image phase shift invariance based cloud motion displacement vector calculation method for ultra-short-term solar PV power forecasting. *Energy Convers. Manag.* **2018**, *157*, 123–135. [CrossRef]

17. Pierro, M.; Bucci, F.; De Felice, M.; Maggioni, E.; Moser, D.; Perotto, A.; Spada, F.; Cornaro, C. Multi-Model Ensemble for day ahead prediction of photovoltaic power generation. *Sol. Energy* **2016**, *134*, 132–146. [[CrossRef](#)]
18. Aburomman, A.A.; Ibne Reaz, M. Bin A novel SVM-kNN-PSO ensemble method for intrusion detection system. *Appl. Soft Comput. J.* **2016**, *38*, 360–372. [[CrossRef](#)]
19. Li, S.; Wang, P.; Goel, L. A novel wavelet-based ensemble method for short-term load forecasting with hybrid neural networks and feature selection. *IEEE Trans. Power Syst.* **2016**, *31*, 1788–1798. [[CrossRef](#)]
20. Bessa, R.J.; Trindade, A.; Silva, C.S.P.; Miranda, V. Probabilistic solar power forecasting in smart grids using distributed information. *Int. J. Electr. Power Energy Syst.* **2015**, *72*, 16–23. [[CrossRef](#)]
21. Iversen, E.B.; Morales, J.M.; Møller, J.K.; Madsen, H. Short-term probabilistic forecasting of wind speed using stochastic differential equations. *Int. J. Forecast.* **2015**, *32*, 981–990. [[CrossRef](#)]
22. Men, Z.; Yee, E.; Lien, F.-S.; Wen, D.; Chen, Y. Short-term wind speed and power forecasting using an ensemble of mixture density neural networks. *Renew. Energy* **2016**, *87*, 203–211. [[CrossRef](#)]
23. Chui, C.K. *Wavelets: A Tutorial in Theory and Applications (Wavelet Analysis and Its Applications)*; Chui, C.K., Ed.; Academic Press: San Diego, CA, USA, 1992.
24. Azimi, R.; Ghofrani, M.; Ghayekhloo, M. A hybrid wind power forecasting model based on data mining and wavelets analysis. *Energy Convers. Manag.* **2016**, *127*, 208–225. [[CrossRef](#)]
25. Feng, X.; Li, Q.; Zhu, Y.; Hou, J.; Jin, L.; Wang, J. Artificial neural networks forecasting of PM<sub>2.5</sub> pollution using air mass trajectory based geographic model and wavelet transformation. *Atmos. Environ.* **2015**, *107*, 118–128. [[CrossRef](#)]
26. Zhu, T.; Wei, H.; Zhao, X.; Zhang, C.; Zhang, K. Clear-sky model for wavelet forecast of direct normal irradiance. *Renew. Energy* **2017**, *104*, 1–8. [[CrossRef](#)]
27. Panapakidis, I.P.; Dagoumas, A.S. Day-ahead natural gas demand forecasting based on the combination of wavelet transform and ANFIS/genetic algorithm/neural network model. *Energy* **2017**, *118*, 231–245. [[CrossRef](#)]
28. Yang, Z.; Ce, L.; Lian, L. Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods. *Appl. Energy* **2017**, *190*, 291–305. [[CrossRef](#)]
29. Tascikaraoglu, A.; Sanandaji, B.M.; Poolla, K.; Varaiya, P. Exploiting sparsity of interconnections in spatio-temporal wind speed forecasting using Wavelet Transform. *Appl. Energy* **2016**, *165*, 735–747. [[CrossRef](#)]
30. Cui, F.; Deng, X.; Shao, H. Short-term wind speed forecasting using the wavelet decomposition and AdaBoost technique in wind farm of East China. *IET Gener. Transm. Distrib.* **2016**, *10*, 2585–2592.
31. Mallat, S.G. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 674–693. [[CrossRef](#)]
32. Meng, A.; Ge, J.; Yin, H.; Chen, S. Wind speed forecasting based on wavelet packet decomposition and artificial neural networks trained by crisscross optimization algorithm. *Energy Convers. Manag.* **2016**, *114*, 75–88. [[CrossRef](#)]
33. Lave, M.; Kleissl, J. Cloud speed impact on solar variability scaling—Application to the wavelet variability model. *Sol. Energy* **2013**, *91*, 11–21. [[CrossRef](#)]
34. US Department of Commerce, NOAA, Earth System Research Laboratory. ESRL Global Monitoring Division—Global Radiation Group. Available online: <https://www.esrl.noaa.gov/gmd/grad/surfrad/> (accessed on 13 September 2017).

