The Ultra-Short-Term Forecasting of Global Horizontal Irradiance Based on Total Sky Images

Junxia Jiang 1,2, Qingquan Lv 3 and Xiaoqing Gao 1,*

1 Key Laboratory of Land Surface Process and Climate Change in Cold and Arid Regions, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China; jiangjx@lzb.ac.cn
2 College of Earth Science, University of Chinese Academy of Sciences, Beijing 100049, China
3 School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China; lvqq18@lzu.edu.cn
* Correspondence: xqgao@lzb.ac.cn

Received: 8 October 2020; Accepted: 5 November 2020; Published: 9 November 2020

Abstract: Solar photovoltaics (PV) has advanced at an unprecedented rate and the global cumulative installed PV capacity is growing exponentially. However, the ability to forecast PV power remains a key technical challenge due to the variability and uncertainty of solar irradiance resulting from the changes of clouds. Ground-based remote sensing with high temporal and spatial resolution may have potential for solar irradiation forecasting, especially under cloudy conditions. To this end, we established two ultra-short-term forecasting models of global horizontal irradiance (GHI) using Ternary Linear Regression (TLR) and Back Propagation Neural Network (BPN), respectively, based on the observation of a ground-based sky imager (TSI-880, Total Sky Imager) and a radiometer at a PV plant in Dunhuang, China. Sky images taken every 1 min (minute) were processed to determine the distribution of clouds with different optical depths (thick, thin) for generating a two-dimensional cloud map. To obtain the forecasted cloud map, the Particle Image Velocity (PIV) method was applied to the two consecutive images and the cloud map was advected to the future. Further, different types of cloud fraction combined with clear sky index derived from the GHI of clear sky conditions were used as the inputs of the two forecasting models. Limited validation on 4 partly cloudy days showed that the average relative root mean square error (rRMSE) of the 4 days ranged from 5% to 36% based on the TLR model and ranged from 12% to 32% based on the BPN model. The forecasting performance of the BPN model was better than the TLR model and the forecasting errors increased with the increase in lead time.

Keywords: solar radiation; solar energy; irradiation; forecasting; image retrieval; Total Sky Imager

1. Introduction

Solar photovoltaics (PV) is developing at an unprecedented speed, which is attributed to dramatic cost reductions, technology advancements and government policy support [1–3]. According to the IEA (International Renewable Energy Agency) report, global solar PV power generation increased by 22% to 720 TWh in 2019 [4]. However, the ability to forecast PV power remains a key technical challenge for the integration of large-scale PV into the electrical grid due to the variability and uncertainty of solar irradiance, which affects PV power most directly [5–9]. Moreover, accurate forecasting of solar irradiance is crucial to the dispatch and management of power systems [5]. Further ultra-short-term systems for forecasting highly spatially and temporally resolved solar irradiance in a timeframe of 15 min have been put forward to optimize the operation of solar power plants [10–12]. In fact, the amount of GHI (global horizontal irradiance) is closely related to atmospheric conditions including cloud, transparency, aerosol concentration and water content [13], in which cloud is the most dominant factor [14,15],
and rapid fluctuations of GHI are generally caused by large changes in cloud conditions [16]. At present, the forecasting of cloud fraction and location is very difficult according to physical and dynamics principles, although it is vital to the ultra-short-term forecasting of solar irradiance [17].

Although well-established methods have been developed for operational solar irradiance forecasting, in the previous studies, part of solar irradiance forecasting applications based on statistical models or artificial neural networks do not directly use characteristic information of cloud, resulting in significant biases [18–22], and the numerical weather prediction (NWP) models that can infer local cloud cover information through the dynamic modeling of atmosphere are widely used for solar irradiance forecasting of several hours to a few days ahead [23,24], whereas they are unlikely to capture real-time temporal relationships between solar irradiance and meteorological factors and simulate cloud with severe convection or short life cycle, due to the coarse spatial resolutions and uncertain initial conditions of NWP models and complicated microscopic physical characteristics of cloud particles [10,25], eventually leading to random errors in short-term solar irradiance prediction [26–28]. Satellite-based solar irradiation forecasting of 1–6 h ahead does not have adequate fine temporal and spatial resolutions to extract information of small-scale and low-altitude cloud [29]. Hence, ground based remote sensing with high temporal and spatial resolution was developed for solar irradiation forecasting to fill the gap of NWP and satellite-based models in nowcasting applications [30,31].

In recent years, sky imagers that rely on ground measurements combined with irradiance sensors have been widely used for nowcasting systems. For Total Sky Imager (TSI), cloud indices derived from images were used for nowcasting of GHI and were reported to have potential for solar irradiance forecasting of 15–20 min ahead [30]. Furthermore, the intra hour forecasting applications for direct normal irradiance (DNI) or GHI of 20 min ahead were performed and the results showed better performance than the persistence model; these studies generally combined the TSI images and the cloud base height for cloud tracking [32–34]. Other studies, in which predictive cloud indices inferred from the ground-based measurements were applied to artificial neural network (ANN) or multilayer-perceptron-network (MLP) models, improved intra hour irradiance forecasting results [35,36]. The radiative transfer model was combined with predicted sky images to forecast GHI and DNI for time horizons of 1–10 min ahead [37]. Recently, minutely solar irradiance was forecasted though the mapping relationship between sky images and irradiance [38]. Moreover, raindrop detection for sky images was proposed to improve the accuracy of short-term solar irradiance [39]. It is obvious that various technologies can be applied to sky images for short-term solar irradiance.

The goal of this paper is to establish two ultra-short-term rolling forecasting models of GHI based on Ternary Linear Regression (TLR) and Back Propagation Neural Network (BPN) using the observation of a sky imager (TSI-880) and GHI in Dunhuang, China. The TLR model was used to establish the real-time temporal linear relationship between three different sky types (thick cloud, thin cloud and clear sky) derived from images and the clear sky index derived from clear sky model, to determine the attenuation proportion of the three sky types to the clear sky index, while the BPN model was used for constructing the nonlinear relation between cloud and clear sky index to forecast GHI.

2. Instruments and Data

The Total Sky Imager of TSI-880 deployed on the rooftop of a solar plant in Dunhuang (set up at 94.478°E, 40.078°N, 1221 m MSL (mean sea level)) was mainly used for ultra-short-term forecasting of the power generation of photovoltaics. The appearance and composition of TSI-880 are shown in Figure 1, consisting of a sun tracking and blocking device, imaging device, and a built-in micro-computer used for the storage and transfer of data. The dome of mirror rotates at a certain speed to ensure that the shadow band blocks strong direct sunlight and prevents the lens from being damaged, and captures the sky scene reflected on the hemisphere mirror with a downward charge coupled device (CCD) camera.

A solar shortwave radiation meter of TBQ-2 was used to measure the instantaneous GHI accurately. The images with a resolution of 352 × 288 pixels and a 24-bit color JPEG format and GHI data were
sampled synchronously with the frequency of 1 min, and simultaneously stored locally and remotely to guarantee the timeliness, integrity and reliability of the data.

![Figure 1. Appearance and composition of total sky imager (TSI-880).](image)

3. Methods and Results

3.1. Cloud Detection and Retrieve

Feature extraction from cloud images is a key component in the application of the Total Sky Imager [12]. Although TSI-880 provides cloud detection according to the built-in algorithm, the single fixed red-blue-Ratio (RBR) threshold for all images and untreated shielding caused significant error in the results. To be specific, on the one hand, the cloud information near the sun shielded by the shadow band is vital for ultra-short-term irradiance forecasting [34]. On the other hand, the image RBR is greatly affected by atmospheric conditions including transparency, aerosol content and sand dust, even though for the same clear sky image, the RBR values around the sun and near the horizon are larger due to aerosol forward scatter and long optical path [40,41]. Given this, we performed cloud detection based on the RBR threshold to determine the cloud fraction.

The first step of cloud detection is recovering the shielding areas that result from the support arm and shadow band, as shown in Figure 2a. The angle of the shadow band deviates from the true north on the image, which is equal to the solar azimuth (SAA), because projection of the sky on the image is isometric [42], and the area of shadow band is shown in the green area of Figure 2b and interpolated according to the pixels on both sides. The projection of the support arm with a fixed position is shown in the white area of Figure 2b and is restored by horizontal bidirectional interpolation. The restored image is shown in Figure 2c. Additionally, the sky, effectively occupying 260 × 260 pixels, is cropped from the original image and distortion is corrected by the spherical distortion correction model [43].

![Figure 2. TSI image: (a) Raw image, (b) occlusion location, (c) occlusion recovery.](image)

In the TSI spherical coordinate system, the relationship of the zenith angle, azimuth and scattering angle (marked with Θ) is shown in Figure 3. The small black square indicates a point on the spherical mirror, θ' and φ' are the zenith angle (IZA) and the azimuth angle of the point, respectively; θ and φ are the solar zenith angle (SZA) and SAA, respectively. The IZA is used to represent the zenith angle of the pixel on the image, and the SPA (Sun-Pixel Angle) is the angle between the sun and the pixel, representing the positional relationship between the pixel and the sun on the image. The SPA
is approximately equal to $\Theta$ due to the long distance between the sun and the earth. From Figure 3, we have:

\[
\sin(IZA) = \sin \theta' = \frac{r}{R}
\]

\[
\cos(SPA) = \cos \Theta = \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos(\phi - \phi')
\]

where $r$ is the distance between pixel and image center on the image, and $R$ is the radius of the image.

![Figure 3. Relationship of the scattering angle, zenith angle and azimuthal angle in spherical coordinates (from ref. [44]).](image)

The images with the SZA of 58° under clear-sky conditions were selected for representing the relationship of the RBR, IZA and SPA, as shown in Figure 4a. It is obvious that the RBR values decrease with the increase in SPA, while they increase with the increase in IZA. The areas of near horizon (IZA > 75°) and around the sun (SPA < 35°) have larger values of RBR in relation to other areas. Hence, the Clear Sky Library (CSL) was applied to eliminate the dependence of RBR on the areas around the sun and horizon [45]. To be specific, the historical clear images were selected as background images to provide RBR reference for images to be detected, and SZA, IZA, SPA, R, G, and B brightness values of historical clear images were saved as the data fields of the CSL. Additionally, we calculated the standard deviation distribution of the RBR of different clear sky images (Figure 4b) and found that the RBR variation had a large range from 0.1 to 0.18 in the area around the sun (SPA < 35°). Given this, the closest image in the CSL was selected for the images to be detected, to reduce the impact of aerosols concentration and sun position projection on the RBR.

![Figure 4. Distribution of RBR (red-blue-ratio) with SPA (Sun-Pixel Angle) and IZA (image zenith angle) in clear sky condition: (a) RBR distribution on 11 August 2015 at SZA = 58° (solar zenith angle); (b) standard deviation of RBR on 4 May, 10 June, 11 August and 24 November 2015 at SZA = 58°.](image)

As for the thresholds of different sky types, a certain number of images taken on different dates during the observation were selected as training samples. The thresholds shown in Figure 5 were derived from the PDF (probability distribution function) of RBR differences of pixels between the selected images of CSL and the sample images, and the thresholds of clear sky (T_clear) and thick cloud (T_thick) were 0.067 and 0.225, respectively.
A pixel was identified as thick cloud if the difference of RBR between the pixel and the corresponding pixel in the background image (ΔRBR) was larger than 0.225; if the ΔRBR was smaller than 0.067, the pixel was identified as clear sky, otherwise the pixel was identified as thin cloud. An instance of cloud detection is shown in Figure 6e. Moreover, the sun parameter (SP) proposed by Pfiste et al. [16] was introduced to correct the type of the pixels around the sun. SP is a dynamic single threshold defined as the RBR average of the pixels around the sun (Figure 4a, −35° < SPA <35°). If the sun is obscured by thick cloud (Figure 6a), RBR of the pixels around the sun are usually smaller than that of the clear day, leading to misidentification as clear sky (Figure 6e). Consequently, when the RBR of the pixel around the sun was larger than the SP, the pixel was re-recognized as thick cloud (Figure 6f). Finally, the cloud fraction of thick cloud ($f_{thick}$) and thin cloud ($f_{thin}$) are the ratios of the numbers of thick cloud and thin cloud to total pixels, respectively.

![Figure 5. Probability distribution of three types of pixel RBR difference on sample images.](image1.png)

![Figure 6. Process of cloud detection (a) image at 15:10 on 7 May 2015 at SZA = 58°, SAA = 239° (solar azimuth). (b) RBR image. (c) The background RBR image at 15:13 on 7 May 7 at SZA = 58°, SAA = 239°. (d) RBR difference of (b,c). (e) Output of cloud detection. (f) Cloud map corrected by SP (sun parameter) (SP = 0.78), white: thick cloud, gray: thin cloud, blue: clear sky.](image2.png)

3.2. Cloud Velocity and Cloud Map Forecasting

Forecasting of the cloud map is a prerequisite for estimating solar irradiance with total sky images. The MATLAB-based algorithm “mpiv (particle image velocimetry)” developed by Mori and Chang...
was applied to the two consecutive images for estimating the translational speed of the cloud [46]. The algorithm has been successfully applied to cloud velocity estimation in solar irradiance forecasting with all-sky images, infrared images and satellite images [32,47,48]. To avoid errors caused by cloud detection, the red channel of the raw images was extracted to calculate cloud speed because “mpiv” can only be used for two-dimensional images, and the red channel of the image has higher contrast to clear sky and cloud than the blue and green channels. The interval of the two consecutive images was 1 min, and the search sub-window was defined as an area of 32 × 32 pixels. Two iterative recursive checks were performed using the MQD (Minimum Quadric Differences) algorithm to calculate the velocity vector field. Median filtering was used to remove discrete error vectors, and finally linear interpolation was used to correct the error vectors.

The cloud velocities were assumed to have the same direction and magnitude due to the limited sight of the TSI. For the clear sky area, the calculated velocities are usually zero or close to zero. Therefore, the calculated velocity vector fields were divided into two categories with the K-means algorithm and the larger value of cluster centers was selected as the representative velocity vector of the cloud. The red and blue scatters represent the velocity fields of the cloud and clear sky, respectively, and the black cross indicates the cluster center in Figure 7. The velocity vector of the cloud was $u = 10.3, \; v = 7.9 \text{ (pixels/min)}$; $u$ and $v$ represent the velocity of the horizontal and vertical direction, respectively.

![Figure 7. Determination of representative velocity vector.](image)

To obtain the forecasted cloud map, the retrieved cloud map was advected to the future with the representative velocity vector. However, the information at the boundary of images is likely to be unpredictable due to the limited observation view of TSI. Here, the boundary of the forecasted cloud map was filled with the boundary information of the image 1 min ago. From Figure 8, we found that the pixels at the edge were more error-forecasted because these areas use the boundary information of the previous moment, and the pixels of thin clouds are also more error-forecasted because the thin clouds are difficult to be accurately detected in the raw image, especially in the area around the sun. The pixels of clear sky are likely to be detected as thin clouds due to the forward scattering of aerosols and haze. Therefore, the accuracy of cloud forecasting is limited by the TSI observation view and cloud detection errors.
The variation of GHI is assumed to be caused by changes in cloud fraction and type (thick cloud and thin cloud) due to the small observation scope. The clear sky index, defined as the ratio of GHI to extraterrestrial radiation ($I_{ets}$) under conditions of clear sky, while it shows random and irregular fluctuations under conditions of broken sky with active clouds. Consequently, many irradiance forecasting models were developed from clear-sky radiation models. The existing clear-sky models such as Solis, Kasten and Ineichen typically developed from atmospheric radiative transfer models can simulate the GHI well, but usually require additional meteorological factors such as aerosol optical depth, ozone content, water vapor content and turbidity [51]. In this study, the clear-sky model was established between the $I_{ets}$ and GHI of clear days during the observation though the least square method.

$$I_{ets} = S_c (1 + 0.033 \times \cos (\frac{360}{365} \times d)) \cos \theta$$

(3)

$S_c$ is the solar constant (1367 W/m²) [49], $d$ is the DOY (day of year) of the year, 1 January is 1 and 31 December is 365.

Diurnal variation of $I_{ets}$ is a cosine function of the zenithal angle [50]. The GHI exhibits uniform attenuation of $I_{ets}$ under conditions of clear sky, while it shows random and irregular fluctuations under conditions of broken sky with active clouds. Consequently, many irradiance forecasting models were developed from clear-sky radiation models. The existing clear-sky models such as Solis, Kasten and Ineichen typically developed from atmospheric radiative transfer models can simulate the GHI well, but usually require additional meteorological factors such as aerosol optical depth, ozone content, water vapor content and turbidity [51]. In this study, the clear-sky model was established between the $I_{ets}$ and GHI of clear days during the observation though the least square method.

$$I_{clk} = \sum_{n=0}^{2} a_n I_{ets}^n, \quad a_n = [-0.00000476, 0.8123113, -35.9099519]$$

(4)

where $I_{clk}$ is the GHI under clear sky, $a_n$ are fitting coefficients.

The three-dimensional radiation transmission process in the clouds and the cloud radiative effect of clouds are very complicated; it is difficult to quantify the forcing effect of clouds on radiation. The variation of GHI is assumed to be caused by changes in cloud fraction and type (thick cloud and thin cloud) due to the small observation scope. The clear sky index, defined as the ratio of GHI to $I_{ets}$, is used to reflect the attenuation degree of entire atmosphere to $I_{ets}$. Whereas in the studies of radiation forecasting, the clear sky index ($k_t$) is usually defined as the ratio of GHI to clear sky GHI [31], meaning that only the effect of the cloud is considered while the influence of other radiative forcing...
factors, including atmospheric molecules and aerosols, is excluded. Additionally, the attenuation degree of radiation by clouds with different optical thicknesses varies considerably. Thick clouds with low-altitude can reduce DNI to almost zero in a few seconds [33], while thin clouds with high-altitude may reduce GHI by 10% [14,52]. Hence, we established GHI estimation models based on the principle of the attenuation degree of cloud to irradiance varying with the distribution proportions of the different types of sky (thick cloud, thin cloud, and clear).

3.4. Forecasting of GHI

3.4.1. Two Forecast Models of GHI

Based on the retrieved and forecasted cloud fraction and clear sky index, the two rolling forecasting models of GHI using ternary linear regression (TLR) and BP neural network were established. Rolling forecasting refers to training the previous model with n values observed from time t-n to t, and the model is used for the forecasting of future time.

The TLR model constructed the relationships of $k_t$ and cloud fraction as shown in Formula (5), describing the attenuation proportion of three sky types (thick cloud, thin cloud and clear sky) 5 min before forecast time, for correcting the clear sky index dynamically.

$$k_{t-5,t} = a_{t-5,t} + b_1(t-5,t)f_{\text{thick}(t-5,t)} + b_2(t-5,t)f_{\text{thin}(t-5,t)} + b_3(t-5,t)f_{\text{clear}(t-5,t)}$$

where $f_{\text{thick}}, f_{\text{thin}}, f_{\text{clear}}$ are the fractions of thick cloud, thin cloud and clear sky. The subscript $t-5,t$ indicates 5 min from $t-5$ to $t$.

The regression coefficients obtained by formula 5 combined with forecasted cloud fraction were applied to estimate the $k_t$ and GHI of the short-term future.

$$k_{t+\Delta t} = a_{t-5,t} + b_1(t-5,t)f_{\text{thick}(t+\Delta t)} + b_2(t-5,t)f_{\text{thin}(t+\Delta t)} + b_3(t-5,t)f_{\text{clear}(t+\Delta t)}$$

where, $\Delta t$ represents the forecasting time scale (taken as 0~5 min in this paper). The subscript $t+\Delta t$ represents the forecasting time.

The BP (Back Propagation) neural network is a multi-layer feed-forward neural network trained by an error back-propagation algorithm which can obtain complex non-linear processing capability through complex mapping of simple non-linear processing elements. Its basic principle is to use the gradient descent method to minimize the mean square error between the actual output and the expected output of the network. It consists of the input layer, the hidden layer, and the output layer. The hidden layer consists of several neural units and each neural unit has a randomly assigned threshold and assigns a weight to the input layer and output layer. The calculation process includes the forward propagation of the signal and the backward propagation of the error. The error output is calculated in the direction from input to output, and the weight and threshold are adjusted in the direction from output to input. In this paper, a BP neural network is used to establish a ultra-short-term GHI forecast model. The training process of the prediction model is simply expressed as Equation (10):

$$(\text{ANN}_u(L,N), I, T) \xrightarrow{\text{training}} \text{ANN}_B$$ (8)

$\text{ANN}_u$ is an untrained BP neural network, $L$ is the number of hidden layers, $N$ is the number of neurons in each hidden layer. $I$ represents inputs including $f_{\text{thick}(t-10,t)}, f_{\text{thin}(t-10,t)}, I_{ets}$, and $I_{lik}$, and $T$ represents the target. $\text{ANN}_B$ is the trained BP neural network. The forecasting model is given by:

$$\hat{I}(t+\Delta t) = \text{ANN}_B(I_t)$$ (9)

where $I_t$ represents new inputs that do not contain the samples involved in training.
Forecasting accuracy and operating efficiency are important indicators for evaluating the performance of the ultra-short-term forecasting model due to the demand of online rolling learning. This paper selected 3, 20, and 10 for $L$, $N$ and the number of samples, respectively, by training the $N$, $L$, the number of iterations and sample error and considering the accuracy of forecasting results and the efficiency of running time.

3.4.2. Evaluation of Forecast Models

Pearson correlation coefficient ($R$) given by formula 10 is used to evaluate the correlation between the forecasted and measured GHI:

$$ R = \frac{\sum_{n=1}^{N} (I_{\text{fh}} - I_{\text{obs}})(I_{\text{obs}} - I_{\text{obs}})}{\sqrt{\sum_{n=1}^{N} (I_{\text{fh}} - I_{\text{obs}})^2} \sqrt{\sum_{n=1}^{N} (I_{\text{obs}} - I_{\text{obs}})^2}} \tag{10} $$

$RMSE$ (Root Mean Square Error) and $rRMSE$ (relative RMSE) are used to indicate the deviation degree of the predicted and measured irradiance, and smaller values for $RMSE$ and $rRMSE$ indicate higher forecasting accuracy and smaller deviation between the forecasted and observed GHI.

$$ RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (I_{\text{fh}} - I_{\text{obs}})^2} \tag{11} $$

$$ rRMSE = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (I_{\text{fh}} - I_{\text{obs}})^2 \times \frac{100}{I_{\text{obs}}} \tag{12} $$

$MAE$ (Mean Absolute Error) is the average of absolute error of the predicted and measured GHI, which can better reflect the prediction error. $MAE$ and $rMAE$ (relative Mean Absolute Error) can be defined by:

$$ MAE = \frac{1}{N} \sum_{n=1}^{N} |I_{\text{fh}} - I_{\text{obs}}| \tag{13} $$

$$ rMAE = \frac{1}{N} \sum_{n=1}^{N} |I_{\text{fh}} - I_{\text{obs}}| \times \frac{100}{I_{\text{obs}}} \tag{14} $$

$MBE$ (Mean Bias Error) is the average value of the predicted and measured GHI errors; $MBE$ and $rMBE$ (relative Mean Bias Error) are given by:

$$ MBE = \frac{1}{N} \sum_{n=1}^{N} (I_{\text{fh}} - I_{\text{obs}}) \tag{15} $$

$$ rMBE = \frac{1}{N} \sum_{n=1}^{N} (I_{\text{fh}} - I_{\text{obs}}) \times \frac{100}{I_{\text{obs}}} \tag{16} $$

3.4.3. Results of GHI Forecasting

Clear or overcast days will always result in perfect forecasting, making them unsuitable for reflecting the ground irradiance response to changes in cloud fraction and evaluating the performance of forecasting models with the sky imager; hence, these days are eliminated. The 4 representative days with partial clouds were selected from November 2017, including 5th November, 7th November, 8th November, and 11th November.

Although nowcasting of GHI is not affected by the cloud motion, it can be used as a verification method for the accuracy of the cloud detection and the significance of the ternary linear regression
equation. Hence, the nowcasting of GHI obtained from formula 5 was used for verifying the rationality of the TLR model in this paper. The results of GHI nowcasting for 5 November 2017 and 7 November 2017 are shown in Figure 9a, and Figure 10a showed that the correlation between forecasted GHI and observed GHI is almost 1, indicating that all the nowcasting GHI can match the observed GHI. Additionally, from Figures 9 and 10b,c, we also found that the forecasted GHI could adequately match the randomness and uncertainty of the observed GHI. In other words, the ternary linear regression equation based on the three different sky types has a significant regression effect, and can consequently be applied for further forecasting.

![Figure 9](image1.png)  
**Figure 9.** Nowcasting irradiance scatter plots on (a) 5 November 2017, (b) 10:00—12:00 on 5 November 2017, (c) 12:00—14:00 on 5 November 2017, (d) 14:00—16:00 on 5 November 2017.

![Figure 10](image2.png)  
**Figure 10.** Nowcasting irradiance scatter plots on (a) 7 November 2017, (b) 10:00—12:00 on 7 November 2017, (c) 12:00—14:00 on 7 November 2017, (d) 14:00—16:00 on 7 November 2017.
The scatter diagrams shown in Figure 11 are the result of GHI forecasting ranging from 1 to 5 min on 5 November 2017 and the various errors are given in Table 1. The points dispersion increases with the increase in time scale, with a validation of $r = 0.999, 0.983, 0.967, 0.948$ for 1 min, 2 min, 3 min, 5 min, respectively. From Table 1, we also found that the forecasting errors increase with the increase in the forecast time. The various errors during the 4 days were given by Tables 1–4. As expected from the previous studies, higher lead times resulted in higher errors. The average RMSE of the 4 selected days for 1 min, 2 min, 3 min, and 5 min forecasting are $39.89 \text{ W/m}^2$, $59.02 \text{ W/m}^2$, $76.64 \text{ W/m}^2$ and $103.08 \text{ W/m}^2$; the average rRMSE are $14\%$, $21\%$, $27\%$, $36\%$, respectively. Bias values (MBE and rMBE) showed negative values, indicating that more sky clear pixels were incorrectly assigned as cloud.

![Figure 11. Scatter plots of global horizontal irradiance (GHI) forecasting based on ternary linear regression (TRL) model for 1 min (a), 2 min (b), 3 min (c), 5 min (d) on 5 November 2017.](image1)

By taking the neural network dynamic rolling learning method, the comparisons between the forecasted and measured GHI on 5 November and 11 November 2017 are given by Figures 12 and 13. The points with larger deviation correspond to the moment when the radiation changes sharply. Bias values (MBE and rMBE) also showing negative values may be related to excessive boundary clouds.

The comparisons of forecasting errors for the 4 days are shown in Tables 1–4. The 4-day average error is shown in Table 5. It is obvious that the errors of RMSE, MAE and MBE of the two kinds of forecasting all increase with the increase in lead time, and the correlation coefficient decreases with the increase in lead time.
(MBE and rMBE) showed negative values, indicating that more sky clear pixels were incorrectly assigned as cloud.

Figure 11. Scatter plots of global horizontal irradiance (GHI) forecasting based on ternary linear regression (TRL) model for 1 min (a), 2 min (b), 3 min (c), 5 min (d) on 5 November 2017.

By taking the neural network dynamic rolling learning method, the comparisons between the forecasted and measured GHI on 5 November and 11 November 2017 are given by Figures 12 and 13. The points with larger deviation correspond to the moment when the radiation changes sharply. Bias values (MBE and rMBE) also showing negative values may be related to excessive boundary clouds.

Figure 12. Scatter plots of GHI forecasting based on the BPN model for 1 min (a), 2 min (b), 3 min (c), 5 min (d) on 5 November 2017.

Figure 13. Scatter plots of GHI forecasting based on the BPN model for 1 min (a), 2 min (b), 3 min (c), 5 min (d) on 11 November 2017.

Although the cloud fraction and clear sky index satisfy the multiple linear regression relation in a short time (several minutes), when using this linear relation to statistically extrapolate the future irradiance, the forecast performance is not as good as the neural network forecast model and the error is larger. The reason for this is that, under broken sky conditions with rapid changes of cloud fraction and speed, the nonlinear BPN-based model can reflect these changes more effectively. This is because neural networks have the ability to learn and can handle complex non-linear problems.

![Fig 11](image1.png)

![Fig 12](image2.png)

![Fig 13](image3.png)
### Table 1. Forecasting error comparison between the ternary linear regression (TRL) model and the backpropagation neural network (BPN) model on 5 November 2017. RMSE: root mean square error, MAE: mean absolute error, MBE: mean bias error; rRMSE: relative root mean square error, rMAE: relative mean absolute error, rMBE: relative mean bias error.

| Model | Time Scale | RMSE | MAE  | MBE   | rRMSE | rMAE | rMBE | R   |
|-------|------------|------|------|-------|-------|------|------|-----|
| TLR   | 0 min      | 8.593| 4.460| 0.370 | 0.038 | 0.020| 0.002| 0.998|
|       | 1 min      | 46.732| 26.479| 0.279 | 0.203 | 0.115| 0.001| 0.95 |
|       | 2 min      | 64.119| 38.595| −1.107| 0.278 | 0.167| −0.005| 0.91 |
|       | 3 min      | 83.427| 49.840| −4.934| 0.347 | 0.208| −0.021| 0.86 |
|       | 5 min      | 117.063| 71.094| −7.650| 0.502 | 0.305| −0.033| 0.83 |
| BPN   | 1 min      | 16.070| 8.649| −0.971| 0.069 | 0.037| −0.004| 0.993|
|       | 2 min      | 27.431| 14.760| −1.206| 0.118 | 0.064| −0.005| 0.98 |
|       | 3 min      | 37.219| 21.471| 0.017 | 0.152 | 0.088| 0.000| 0.966|
|       | 5 min      | 49.500| 28.701| −1.932| 0.212 | 0.123| −0.008| 0.936|
|       | 10 min     | 73.136| 46.970| −4.762| 0.311 | 0.200| −0.020| 0.864|

### Table 2. Forecasting error comparison between the linear regression model and the neural network model on 7 November 2017.

| Model | Time Scale | RMSE | MAE  | MBE   | rRMSE | rMAE | rMBE | R   |
|-------|------------|------|------|-------|-------|------|------|-----|
| TLR   | 0 min      | 33.320| 15.906| 2.411 | 0.117 | 0.056| 0.008| 0.996|
|       | 1 min      | 48.784| 30.077| 1.221 | 0.169 | 0.104| 0.004| 0.931|
|       | 2 min      | 65.655| 44.577| 0.628 | 0.227 | 0.154| 0.002| 0.882|
|       | 3 min      | 102.039| 68.285| −1.546| 0.352 | 0.236| −0.005| 0.757|
|       | 5 min      | 124.722| 85.332| −3.327| 0.429 | 0.294| −0.011| 0.672|
| BPN   | 1 min      | 43.307| 26.267| 0.065 | 0.151 | 0.092| 0.000| 0.937|
|       | 2 min      | 56.867| 36.142| −2.368| 0.198 | 0.126| −0.008| 0.893|
|       | 3 min      | 81.953| 54.740| −0.672| 0.286 | 0.191| −0.002| 0.801|
|       | 5 min      | 97.350| 63.719| 4.313 | 0.338 | 0.221| 0.015| 0.742|
|       | 10 min     | 124.594| 89.594| 2.770 | 0.430 | 0.309| 0.010| 0.555|

### Table 3. Forecasting error comparison between the linear regression model and the neural network model on 8 November 2017.

| Model | Time Scale | RMSE | MAE  | MBE   | rRMSE | rMAE | rMBE | R   |
|-------|------------|------|------|-------|-------|------|------|-----|
| TLR   | 0 min      | 6.562| 1.885| −0.076| 0.020 | 0.006| 0.000| 0.996|
|       | 1 min      | 32.690| 14.152| 0.958 | 0.096 | 0.042| 0.003| 0.931|
|       | 2 min      | 54.270| 24.522| 2.736 | 0.164 | 0.074| 0.008| 0.882|
|       | 3 min      | 49.692| 24.503| −4.380| 0.149 | 0.074| −0.013| 0.757|
|       | 5 min      | 58.799| 30.687| −6.170| 0.176 | 0.092| −0.018| 0.672|
| BPN   | 1 min      | 47.604| 27.184| 0.368 | 0.173 | 0.099| 0.001| 0.944|
|       | 2 min      | 66.083| 40.428| −5.953| 0.240 | 0.147| −0.022| 0.889|
|       | 3 min      | 81.247| 52.912| −4.567| 0.294 | 0.192| −0.017| 0.841|
|       | 5 min      | 98.843| 65.099| −2.602| 0.357 | 0.235| −0.009| 0.773|
|       | 10 min     | 109.479| 74.403| −5.503| 0.452 | 0.307| −0.023| 0.727|
Table 4. Forecasting error comparison between the linear regression model and the neural network model on 11 November 2017.

| Model | Time Scale | RMSE  | MAE   | MBE   | rRMSE | rMAE  | rMBE  | R      |
|-------|------------|-------|-------|-------|-------|-------|-------|--------|
| TLR   | 0 min      | 8.579 | 2.973 | −1.494 | 0.026 | 0.009 | −0.005 | 0.999  |
|       | 1 min      | 31.334| 13.912| −4.006 | 0.092 | 0.041 | −0.012 | 0.988  |
|       | 2 min      | 52.052| 22.725| −7.152 | 0.157 | 0.068 | −0.022 | 0.967  |
|       | 3 min      | 71.395| 31.077| −10.971| 0.215 | 0.093 | −0.033 | 0.941  |
|       | 5 min      | 111.740| 45.512| −21.055| 0.335 | 0.136 | −0.063 | 0.875  |
| BPN   | 1 min      | 38.779| 18.536| 1.195  | 0.16  | 0.055 | 0.004  | 0.980  |
|       | 2 min      | 53.112| 26.449| 0.922  | 0.158 | 0.079 | 0.003  | 0.963  |
|       | 3 min      | 55.241| 27.653| −1.22  | 0.165 | 0.083 | 0.006  | 0.961  |
|       | 5 min      | 57.718| 30.609| 0.817  | 0.172 | 0.091 | 0.002  | 0.956  |
|       | 10 min     | 81.365| 50.763| 1.780  | 0.321 | 0.401 | 0.010  | 0.895  |

Table 5. The average forecasting error of 4 days comparison between the linear regression model and the neural network model.

| Model | Time Scale | RMSE  | MAE   | MBE   | rRMSE | rMAE  | rMBE  |
|-------|------------|-------|-------|-------|-------|-------|-------|
| TLR   | 0 min      | 14.26 | 6.31  | 0.30  | 0.05  | 0.023 | 0.001 |
|       | 1 min      | 39.89 | 21.16 | −0.39 | 0.14  | 0.076 | −0.001|
|       | 2 min      | 59.02 | 32.60 | −1.22 | 0.21  | 0.116 | −0.004|
|       | 3 min      | 76.64 | 43.43 | −5.46 | 0.27  | 0.15  | −0.018|
|       | 5 min      | 103.08| 58.16 | −9.55 | 0.36  | 0.21  | −0.031|
| BPN   | 1 min      | 38.78 | 18.54 | 1.20  | 0.12  | 0.055 | 0.004 |
|       | 2 min      | 53.11 | 26.45 | 0.92  | 0.16  | 0.079 | 0.003 |
|       | 3 min      | 55.24 | 27.65 | 2.20  | 0.17  | 0.083 | 0.006 |
|       | 5 min      | 57.72 | 30.61 | 0.82  | 0.17  | 0.091 | 0.002 |
|       | 10 min     | 81.37 | 50.76 | 1.78  | 0.32  | 0.401 | 0.010 |

4. Discussion

The two GHI forecasting models developed in this paper are mainly based on the ground observational characteristics of clouds from the Dunhuang region of China, considering the attenuation effect of clouds with different thickness and proportion on GHI. To apply them to other areas, cloud detection and forecasting technologies are universal and necessary adjustment based on observations for the clear sky model are needed. Meanwhile, the radiation transmission in cloudy atmosphere is an extremely complicated process and the interaction mechanism between radiation and clouds remains poorly understood. In fact, clouds with different heights have diverse attenuation effects on radiation. However, cloud base height is not available for a single TSI site. Unified processing of clouds of different heights is likely to increase the error of forecasting.

Moreover, in this study, the temporal–spatial range is limited by the observation range of the ground sky imager. The boundary information of forecasted cloud images comes from the original images and cannot represent the sky situation in the future. When the clouds move fast, the forecasting error of the cloud image increases and the forecasting time scale reduces.

Overall, the average rRMSE of the 4 selected days ranged from 5% to 36% based on the TLR model and ranged from 12% to 32% based on the BPN model. In the previous study, the forecasting with four all-sky cameras, which were claimed to be state of the art, reduced the rRMSE of GHI from 30.9% to 23.5% [11]; however, the computational efficiency and cost were also much higher.

Dunhuang is located in northwestern China; the weather in winter is usually dry. There is almost no typical synoptic process during the observation period, broken days are few and the optical thickness
of cloud is small. On the one hand, thin high-altitude cloud is hard to be accurately detected in the sky images, and the error of cloud detection and retrieval limited the accuracy of irradiance forecasting. On the other hand, thin clouds are prone to move and dissipate quickly, making forecasting more difficult and the time scale smaller.

In view of the shortcomings in the current work, the prospects for future work are proposed: The localized network of TSI observation will be deployed to retrieve the cloud base height and extract the cloud information and join the multiple images from different sites at the same time, achieving the monitoring of cloud motion in a larger area and improving spatial-temporal scale and accuracy of refined forecasting of solar radiation.

5. Conclusions

Two rolling forecasting models of GHI were established in this study based on the observation of a TSI and radiometer at a PV plant in Dunhuang, China. The main technical work included cloud detection and forecasting. To our knowledge, this is first time that such models have been developed and compared. Although the complicated effect of clouds on radiation is simplified in the models, especially in the TLR model, by including the objective observational characteristics of clouds into the surface radiation based on clear sky radiation, we simulated the GHI under the partly cloudy sky conditions. The two models allow us to effectively model the decreasing or increasing trend and degree of local GHI from the deployment of the ground observation of clouds; this is of significant relevance to the scheduling, dispatching and regulation of PV plants’ power.

For cloud detection, the area around the sun and near-horizontal area are difficult to distinguish accurately; the error of thin cloud, especially in the areas around the sun and the shading areas, is greater than thick cloud. The threshold of clear sky and thick cloud is 0.067 and 0.225, respectively, in this study.

The accuracy of cloud map prediction is limited by the accuracy of cloud recognition. Because of the limited view of TSI, there will be clouds moving outside of the image, causing errors in the forecasted cloud map, through advection transport, to increase with the increase in lead time, and causing the average rRMSE of the 4 selected days to range from 5% to 36% based on the TLR model and from 12% to 32% based on the BPN model.

Author Contributions: Conceptualization, J.J.; methodology, J.J.; software, J.J. and Q.L.; validation, Q.L.; writing—original draft preparation, J.J.; writing—review and editing, J.J. and Q.L.; funding acquisition, X.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work is financially supported by the National Key Research and Development Program of China (2018YFB1502800), the National Natural Science Foundation of China (41805085).

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

References

1. Haegel, N.M.; Margolis, R.; Buonassisi, T.; Feldman, D.; Froitzheim, A.; Garabedian, R.; Green, M.; Glunz, S.; Henning, H.-M.; Holder, B.; et al. Terawatt-scale photovoltaics: Trajectories and challenges. *Science* 2017, 356, 141–143. [CrossRef] [PubMed]
2. Breyer, C.; Bogdanov, D.; Gulagi, A.; Aghahosseini, A.; Barbosa, L.S.; Koskinen, O.; Barasa, M.; Caldera, U.; Afanasyeva, S.; Child, M.; et al. On the role of solar photovoltaics in global energy transition scenarios. *Prog. Photovolt. Res. Appl.* 2017, 25, 727–745. [CrossRef]
3. Wilson, G.M.; Al-Jassim, M.; Metzger, W.K.; Glunz, S.W.; Verlinden, P.; Xiong, G.; Mansfield, L.M.; Stanbery, B.J.; Zhu, K.; Yan, Y.; et al. The 2020 photovoltaic technologies roadmap. *J. Phys. D Appl. Phys.* 2020, 53, 493001. [CrossRef]
4. IEA. 2020. Available online: https://www.iea.org/reports/solar-pv (accessed on 1 October 2020).
5. Lave, M.; Kleissl, J. Solar variability of four sites across the state of Colorado. *Renew. Energy* 2010, 35, 2867–2873. [CrossRef]
30. Chow, C.W.; Urquhart, B.; Lave, M.; Dominguez, A.; Kleissl, J.; Shields, J.; Washom, B. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. Sol. Energy 2011, 85, 2881–2893. [CrossRef]

31. Inman, R.H.; Pedro, H.T.C.; Coimbra, C.F.M. Solar forecasting methods for renewable energy integration. Prog. Energy Combust. Sci. 2013, 39, 535–576. [CrossRef]

32. Marquez, R.; Coimbra, C.F.M. Intra-hour DNI forecasting based on cloud tracking image analysis. Sol. Energy 2013, 91, 327–336. [CrossRef]

33. Quesada-Ruiz, S.; Chu, Y.; Tovar-Pescador, J.; Pedro, H.T.C.; Coimbra, C.F.M. Cloud-tracking methodology for intra-hour DNI forecasting. Sol. Energy 2014, 102, 267–275. [CrossRef]

34. Yang, H.D.; Kurtz, B.; Nguyen, D.; Urquhart, B.; Chow, C.W.; Ghonima, M.; Kleissl, J. Solar irradiance forecasting using a ground-based sky imager developed at UC San Diego. Sol. Energy 2014, 103, 502–524. [CrossRef]

35. Marquez, R.; Gueorguiev, V.G.; Coimbra, C.F.M. Forecasting of Global Horizontal Irradiance Using Sky Cover Indices. J. Sol. Energy Eng. Trans.-ASME 2013, 135, 5. [CrossRef]

36. Chu, Y.H.; Li, M.Y.; Coimbra, C.F.M. Sun-tracking imaging system for intra-hour DNI forecasts. Renew. Energy 2016, 96, 792–799. [CrossRef]

37. Du, J.; Min, Q.; Zhang, P.; Guo, J.; Yang, J.; Yin, B. Short-Term Solar Irradiance Forecasts Using Sky Images and Radiative Transfer Model. Energies 2018, 11, 1107. [CrossRef]

38. Wang, F.; Xuan, Z.; Zhen, Z.; Li, Y.; Li, K.; Zhao, L.; Shafie-Khah, M.; Catalaio, J.P. A minutely solar irradiance forecasting method based on real-time sky image-irradiance mapping model. Energy Convers. Manag. 2020, 220, 113075. [CrossRef]

39. Voronych, O.; Höller, R.; Beck, G.L.; Traunmüller, W. Solar PV nowcasting based on skycamera observations. Adv. Sci. Res. 2019, 16, 7–10. [CrossRef]

40. Cantor, A. Optics of the atmosphere–Scattering by molecules and particles. IEEE J. Quantum Electron. 1978, 14, 698–699. [CrossRef]

41. Schade, N.H.; Macke, A.; Sandmann, H.; Stick, C. Enhanced solar global irradiance during cloudy sky conditions. Meteorol. Z. 2007, 16, 295–303. [CrossRef]

42. Huo, J.; Lx, D. A method to retrieve the radiance from all-sky visible images. Acta Meteorol. Sin. 2010, 68, 800–807.

43. Han, Y.; Song, J. An equivalent surface model of image distortion and the correction algorithm. Opt. Technol. 2005, 31, 122–124.

44. Liou, K-N. An Introduction to Atmospheric Radiation, 2nd ed.; Academic Press: San Diego, CA, USA, 2004; p. 259.

45. Shields, J.E.; Johnson, R.W.; Koehler, T.L. Automated Whole Sky Imaging Systems for Cloud Field Assessment; Fourth Symposium on Global Change Studies: Boston, MA, USA; American Meteorological Society: Boston, MA, USA, 1993; pp. 228–231.

46. Mori, N.; Chang, K.A. Introduction to MPIV. Available online: http://www.oceanwave.jp/softwares/mpiv (accessed on 8 October 2020).

47. Chu, Y.H.; Pedro, H.T.C.; Coimbra, C.F.M. Hybrid intra-hour DNI forecasts with sky image processing enhanced by stochastic learning. Sol. Energy 2013, 98, 592–603. [CrossRef]

48. Marquez, R.; Pedro, H.T.C.; Coimbra, C.F.M. Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs. Sol. Energy 2013, 92, 176–188. [CrossRef]

49. Sheng, P.; Mao, J.; Li, J.; Sang, J.; Pan, N. Atmospheric Physics; China Meteorological Press: Beijing, China, 2006; pp. 84–98.

50. Cionco, R.G.; Soon, W.W-H. Short-term orbital forcing: A quasi-review and a reappraisal of realistic boundary conditions for climate modeling. Earth-Sci. Rev. 2017, 166, 206–222. [CrossRef]

51. Ineichen, P. Comparison of eight clear sky broadband models against 16 independent data banks. Sol. Energy 2006, 80, 468–478. [CrossRef]

52. Kasten, F.; Czeplak, G. Solar and terrestrial radiation dependent on the amount and type of cloud. Sol. Energy 1980, 24, 177–189. [CrossRef]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.