

Research Article

Autoregressive Neural Network for Cloud Concentration Forecast from Hemispheric Sky Images

Cristian Crisosto 

Leibniz Universität Hannover, Institute for Meteorology and Climatology, Herrenhäuser Straße 2, 30419 Hannover, Germany

Correspondence should be addressed to Cristian Crisosto; crcrisos@gmail.com

Received 23 January 2019; Revised 15 March 2019; Accepted 20 March 2019; Published 24 April 2019

Academic Editor: Philippe Poggi

Copyright © 2019 Cristian Crisosto. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We present here a new method to predict cloud concentration five minutes in advance from all-sky images using the Artificial Neural Networks (ANN). An autoregressive neural network with backpropagation (Ar-BP) was created and trained with four years of all-sky images as inputs. The pictures were taken with a hemispheric sky imager fixed on the roof at the Institute of Meteorology and Climatology (IMUK) of the Leibniz Universität Hannover, Hannover, Germany. Firstly, a statistical method is presented to obtain key information of the pictures. Secondly, a new image-processing algorithm is suggested to optimize the cloud detection process starting with the Haze Index. Finally, the cloud concentration five minutes in advance at the IMUK is forecasted using machine learning methods. A persistence model forecast to provide a reference for comparison was generated. The results are quantified in terms of the root mean square error (RMSE) and the mean absolute error (MAE). The new algorithm reduced both the RMSE and the MAE of the prediction by approximately 30% compared to the reference persistence model under diverse cloud conditions. The new algorithm could be used as a tool for the stable maintenance of the network for the transmission system operators, i.e., the primary control reserve (within 30 seconds) and the secondary control reserve (within 5 minutes).

1. Introduction

Changes in the solar irradiance dynamic are significantly impacted by clouds, which makes it difficult to achieve accurate PV power forecasting [1, 2]. With a reliable cloud performance forecast, uncertainty in the solar irradiance prediction can be minimized and optimized. Increased electricity demand requires balancing energy. In this case, the grid operator needs additional power supply to his grid at short notice. The prequalification requirements in German markets provide that a complete deployment of primary control reserve has been completed within 30 seconds and the secondary control energy must be available in full within five minutes [3]. Quaschning [4] shows the importance of the energy control for the first 15 seconds, 30 seconds, 1 minute, 5 minutes, 15 minutes, and 60 minutes when, for example, the current reserve is running low. Thus, a prediction of the cloud concentration for the next five minutes, for solar radiation forecast, makes an important contribution to the

efficient and economical application for many areas of solar energy use. Solar irradiance is the key factor for solar photovoltaic (PV) generation. The International Energy Agency estimates that after 2060, solar energy could cover up to a third of the world's energy consumption. Therefore, the solar energy use is likely to grow by a double digit rate throughout the world and for decades.

Thus, solar power will be an important contributor to the future power supply technologies, influencing the planning, profitability, and operation of power systems. For stabilizing the fluctuations in the energy output of PV plants, the impact of clouds must be considered to achieve a sustainable, affordable, and reliable electricity supply [5].

All-sky images have already been proven to achieve an efficient observation from the ground delivering a comprehensive view for kilometers [6]. The application of this technique is mainly used in the solar forecasting for cloud identification, cloud movement, and cloud forecast [7]. Many authors use algorithms based on a red/blue threshold of the

RGB channels of all-sky images for cloud classification [8]. However, this method is unable to detect thin clouds near the horizon [9].

Additionally, several studies report different ways for identifying clouds and predicting their future movements in a more objective manner. Many authors report methods in cloud identification using a threshold and segmentation of the pictures. Liu et al. [10] developed an automatic cloud detection algorithm using superpixel segmentation calculating the local threshold for each superpixel and then determining the threshold matrix for whole images. Scolari et al. [11] developed a cloud motion identification algorithm based on all-sky images for prediction horizons in the range 1 to 10 minutes.

More recently, Crisosto et al. [12] developed an algorithm to predict the global horizontal irradiance (GHI) one hour in advance from all-sky images using the ANN. This study reduced both the RMSE and the MAE of the one-hour prediction by approximately 40% of the forecast prediction compared to the reference persistence model subdividing the all-sky images into concentric circles to be able to simulate more accurately the GHI.

Different methodologies utilizing two ANN have already been employed. Kamadinata et al. [13] developed and compared two different ANN to first forecast cloud movement direction where the output of this ANN is utilized as input for the second ANN for predicting the GHI. The results of this study show a reduction of the computational effort capturing the trend of the GHI very well. Zhen et al. [14] proposed a cloud image forecasting method from all-sky images using genetic algorithms tracking both the displacement and deformation of cloud reducing the Euclidean distance in comparison with other methods.

Therefore, in order to support accurate solar irradiance forecasts, we propose a cloud concentration forecast algorithm using the artificial neural networks (ANN), which can be later used as a tool for solar energy forecasts. Section 2 briefly describes the data and image acquisition. Section 3 describes the methodologies necessary for this study. The forecasting results are given in Section 4. Finally, in section 5, the conclusions and future work will be discussed.

2. Data

The main component of solar power output is the solar irradiance, which under the presence of clouds is extremely affected. Thus, cloud motion becomes the key element of solar power output.

2.1. Image Acquisition. The camera system is installed inside a weatherproof housing on the roof of the Institute of Meteorology and Climatology (IMUK) of the Leibniz Universität Hannover, Hannover, Germany. The pictures were recorded with a Canon EOS 700D equipped with a Dörr DHG fisheye lens providing a 183° field of view. The exposure time of the pictures was 1000/s. All time hours are expressed in coordinated universal time (UTC).

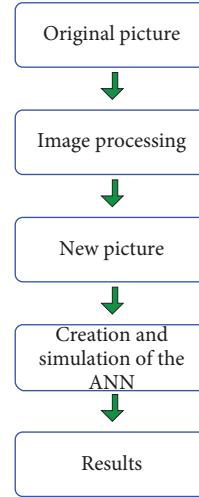


FIGURE 1: Cloud movement process with ANN.

3. Methodology

We developed an algorithm to forecast cloud concentration five minutes in advance from all-sky images using the ANN. We can divide the new algorithm in two main steps. The first part highlights the image-processing algorithm for extracting parameters from all-sky images. The second step comprises the ANN method. Figure 1 shows the cloud recognition and cloud concentration forecast with ANN.

3.1. Cloud Pixel Identification. Clear sky is characterized by high blue pixel intensity and low red pixel intensity, while thick cloud pixels are characterized by high intensity in both channels. Thus, the cloud identification algorithm determines if a pixel corresponds to a cloudy point or clear sky. To surpass the limitations of using only the blue and red RGB channels, we used a cloud identification method which also uses green. This color discrimination method is simple and distinguishes cloud from blue sky by the ratio of the counts of red, green, and blue color in each pixel. Using the Sky Index (equation 1) method by Yamashita et al. [15] and refining its uncertainties, we calculated the Haze Index (equation 2) as detailed by Schrempf [16] to expand (1) for a better cloud identification.

$$\text{Sky Index} = \frac{\text{count}_{\text{blue}} - \text{count}_{\text{red}}}{\text{count}_{\text{blue}} + \text{count}_{\text{red}}}, \quad (1)$$

$$\text{Haze Index} = \frac{((\text{count}_{\text{red}} + \text{count}_{\text{blue}})/2) - \text{count}_{\text{green}}}{((\text{count}_{\text{red}} + \text{count}_{\text{blue}})/2) + \text{count}_{\text{green}}}. \quad (2)$$

The total cloud area was then calculated by the separation of cloud and sky done by the Haze Index in the all-sky image. To avoid oversaturated pixels, the percentage of clear sky and cloud cover is obtained without considering the sun's circumference. The extraction of the

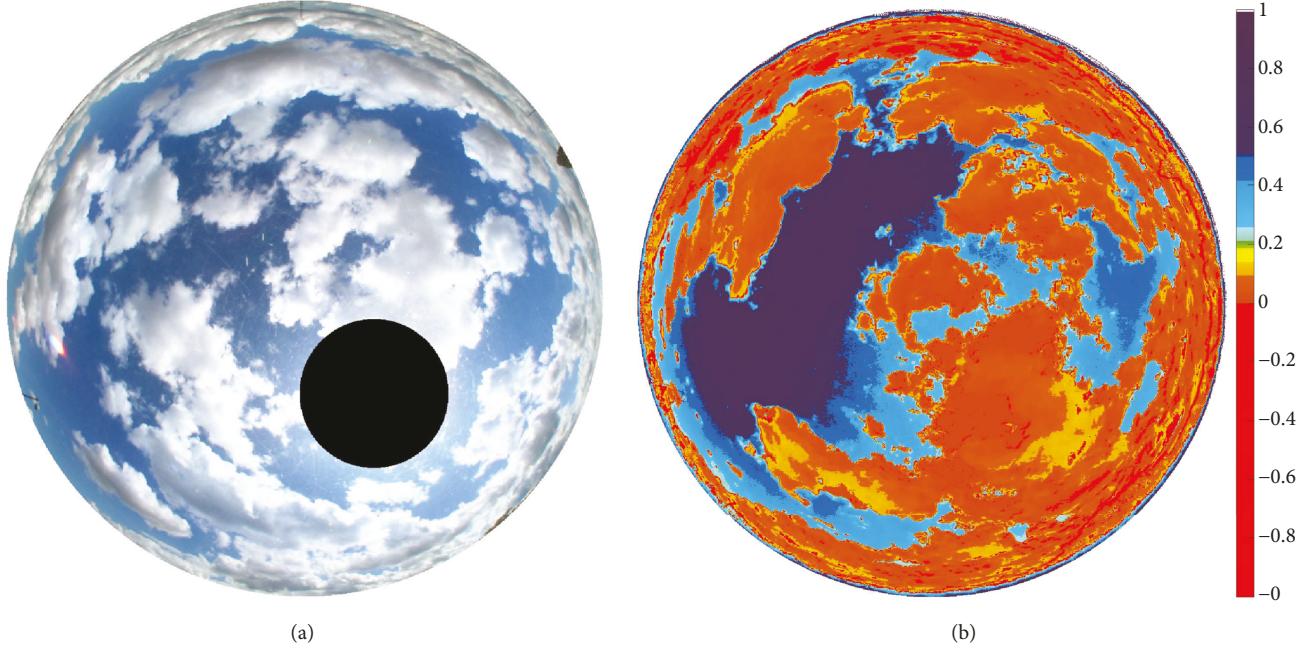


FIGURE 2: (a) Cropped black area and coverage of the sun from the original picture. (b) Haze Index image.

statistical information from all pictures was limited to the sun's zenith angle of 70°. Figure 2 shows the Haze Index image processing.

3.2. Setup of the ANN. The algorithm used seven inputs and one output. For the final configuration of the ANN see Table 1. The inputs x_j flow through the next layer multiplying their values by a weight $w_{i,j}$, while the resulting product is used as argument for a transfer function f giving the output y_i . i represents the presynaptic neuron and j the postsynaptic neuron, see equation 3 and equation 4. The quantity of hidden neurons per single hidden layer was calculated by (5).

$$u_i = \left(\sum_{j=1}^n w_{i,j} x_j \right), \quad (3)$$

$$y_i = f(u_i + b_i), \quad (4)$$

where x_j is the input, w_i is the synaptic weight, u_i is the linear combination of the inputs, b_i is the bias, f is the activation function, and y_i is the output.

$$m = \sqrt{n+l} \pm \alpha, \quad (5)$$

where n is the number of inputs and l is the number of output neurons. α is a constant ($1 < \alpha < 10$).

The build of neural networks has shown that in addition to the number of neurons and layers, the configuration of the initially chosen weights has a significant influence on the network. Thus, different neural network structures were configured to carry out this job. Each one differs in its construction form, initial values, and learning algorithm.

TABLE 1: The neural network structure occupied in this investigation.

Input	Hidden layers	Neurons in the first hidden layer	Neurons in the second hidden layer	Output
7	2	4	2	1

The selection of the “best” network was extremely difficult. For this reason, the technique used in all the networks built was to add hidden layers and in these to add neurons. The idea is to achieve the desired result with the least possible number of hidden layers and neurons in each of these layers.

The interaction of the data decides the quality of the test. The criteria for choosing a network was that both the RMSE and the MAE were minimized as much as possible.

3.3. Cloud Concentration Forecasting. To accurately follow cloud concentrations five minutes in advance, we created the ring program, see Figure 3. The ring program divides the pictures in concentric rings with the sun as their center. We can see in the picture the subdivision of n concentric rings. Each of these rings represents a temporal resolution. The width depends on the distance from the horizon to the center of the sun due to the equidistant projection. Figure 3(a) shows the number of circles on 22nd June 2014 at 12:51 over the original picture, and Figure 3(b) shows the rings at the same time over the Haze Index image. The number of circles in this moment of the day was $n = 10$ that corresponds to approximately 10 minutes of future information, i.e., the time that the clouds could take to reach the center of the sun. In addition, the wind speed is also measured at the IMUK, and each picture is stored with the

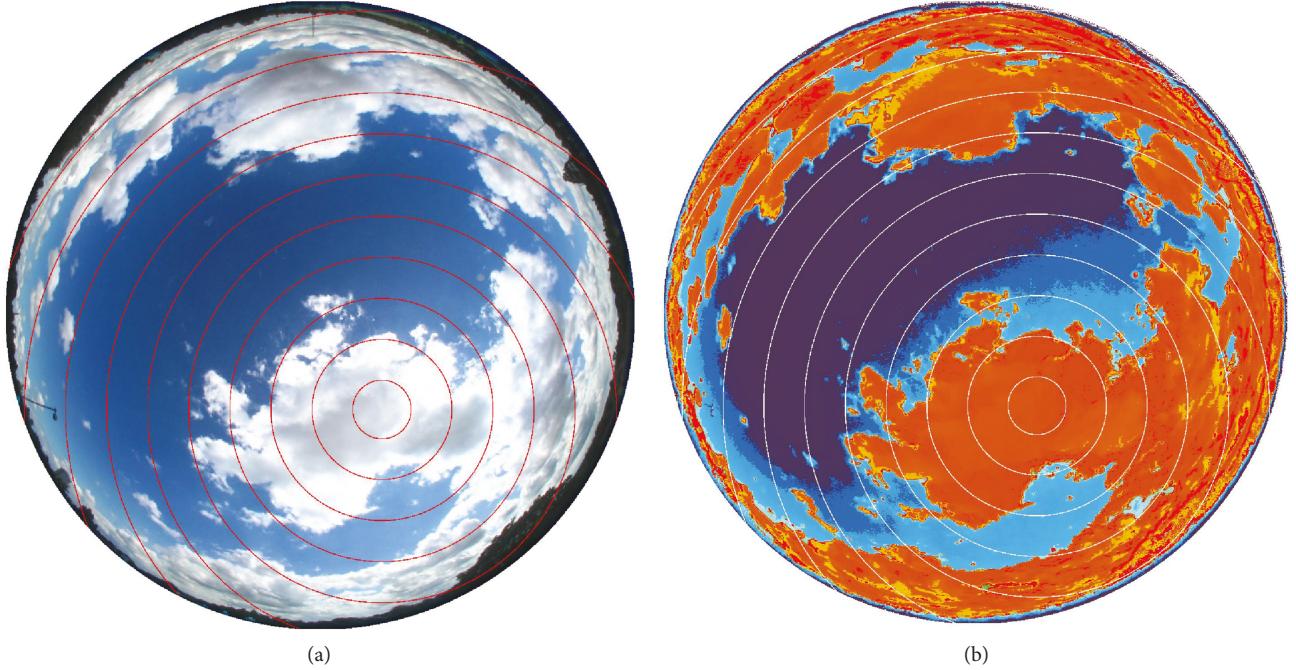


FIGURE 3: (a) shows the 10 circles on 22nd June 2014 at 12:51 on the original image. (b) shows 10 circles on 22nd June 2014 at 12:51 on the Haze Index image.

corresponding wind speed to estimate how many circles there should be in each picture.

In the training phase of the ANN, we used the percentage of cloud in each ring, the sun zenith angle (SZA) and the mathematical standard deviation, mode, median, and average of the RGB channels of each ring at the time t as input parameters. One-minute ahead cloud cover fraction of the next ring at $t + 1$ of the next picture is the output of the program. Now, to predict the cloud concentration at time $t + 2$, we take all inputs from the time t with exception of the cloud cover; this input is taken at time $t + 1$.

For example, on 22nd June 2014, the simulation started at 12:51 (t) and estimated the cloud cover fraction of the next ring of the next picture at 12:52 ($t + 1$) (Figures 4(a) and 4(b)). Subsequently, to forecast the cloud cover fraction of the next ring at 12:53 ($t + 2$), we used all input parameters of the circle at 12:51 (t) with exception of the input cloud cover, which was taken from the forecasted cloud cover of 12:52 ($t + 1$). The idea is to use the information of only one picture to forecast the cloud cover from 1-5 minutes ahead, completing all 5 rings.

Therefore, the ANN analyzed the actual cloud concentration at the current ring in order to know if one minute in advance, the next ring will have the same, larger, or smaller cloud concentration. This information could be important to know the most likely cloud concentration near the sun at the next minutes, in order to know how variable the solar irradiance will be in this time frame.

4. Results

To evaluate the proposed method, the first five minutes from 50 images with different cloud concentrations and

sun positions were manually selected and analyzed. The selected days represent high cloud variability, i.e., a high variability of solar irradiance. The RMSE (6), the MAE (7), and the coefficient of determination (R^2) (8) were used to evaluate the performance of the new model for these five minutes. To finally validate our model, the statistical sampling (9) was utilized and the results are presented as a boxplot. The mathematical definitions of the statistical procedure are expressed as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N \|y_i - x_i\|}{N}}, \quad (6)$$

$$\text{MAE} = \frac{\sum_{i=1}^N |y_i - x_i|}{N}, \quad (7)$$

$$R^2 = \frac{\sum_{i=1}^N (y_i - \bar{y})(x_i - \bar{x})}{\left[\left(\sum_{i=1}^N (y_i - \bar{y})^2 \right) \left(\sum_{i=1}^N (x_i - \bar{x})^2 \right) \right]^{1/2}}, \quad (8)$$

where y_i was the forecast value, x_i was the measured value, and N was the number of samples. Additionally, $\bar{x} = \sum_{i=1}^N x_i$ and $\bar{y} = \sum_{i=1}^N y_i$.

$$x = \frac{Z^2 N p q}{e^2 (N-1) Z^2 p q}, \quad (9)$$

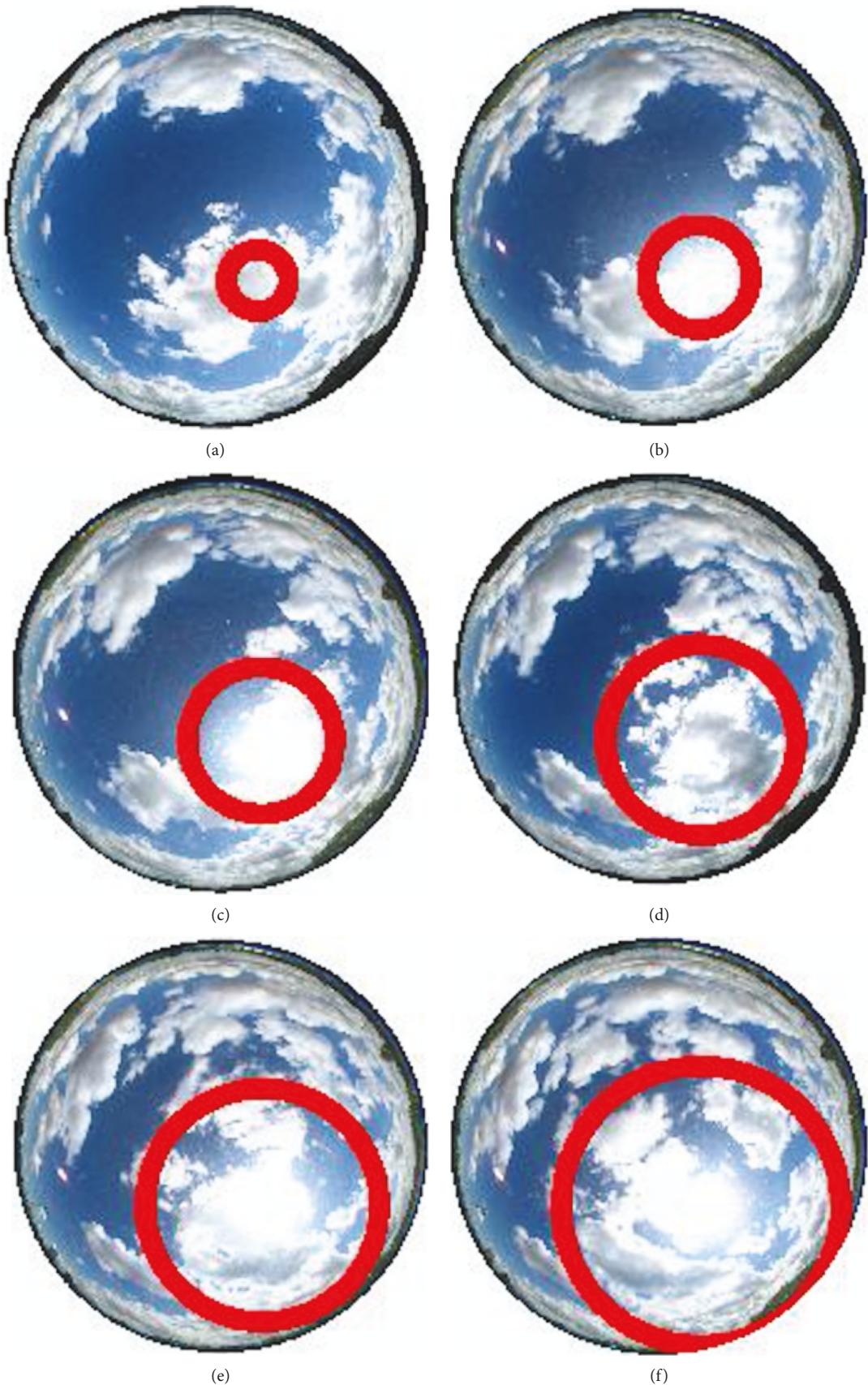


FIGURE 4: Observation of all-sky images on 22nd June 2014. The red circle in (a) corresponds at 12:51. (b) at 12:52. (c) at 12:53. (d) at 12:54. (e) at 12:55. (f) at 12:56.

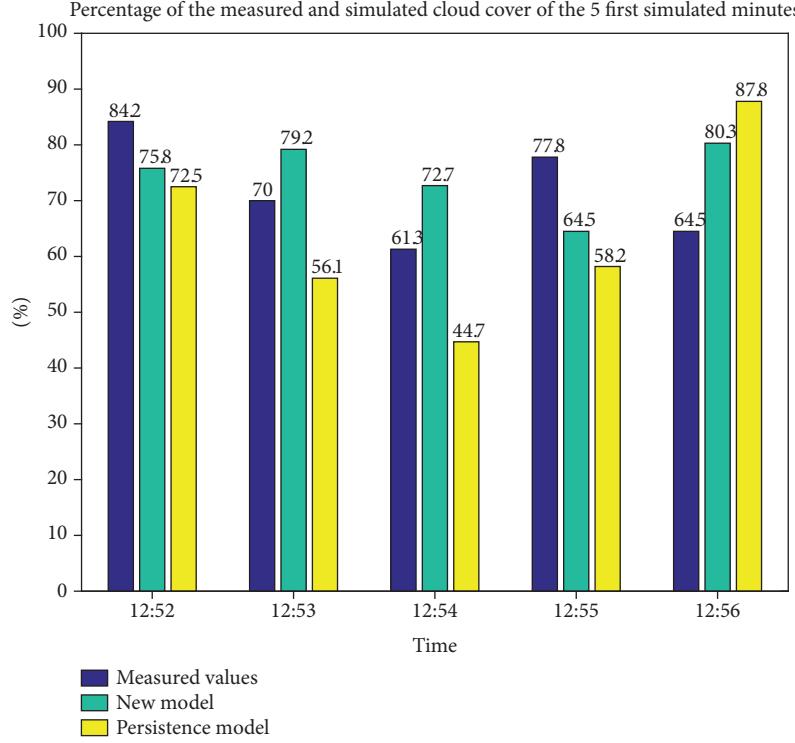


FIGURE 5: Comparison between the measured and simulated cloud cover on 22nd June 2014 from 12:52 until 12:56.

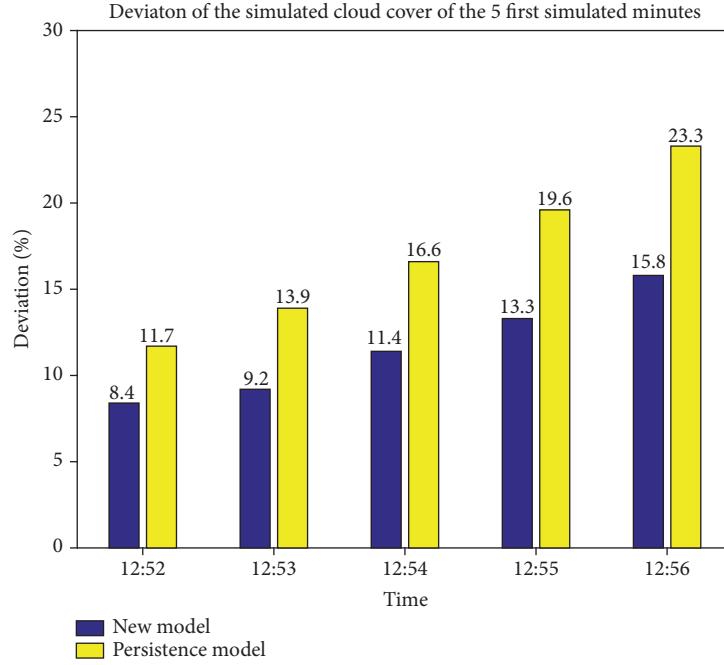


FIGURE 6: Deviation between the simulated cloud cover of the new model and the persistence model on 22nd June 2014 from 12:52 until 12:56.

where N was the total of the set, $p = 0.95$, $q = 0.05$, and $Z = 1.96$ (this value corresponds to the confidence level of 95%).

The new algorithm was compared with the benchmark algorithm, the persistence model. This model is the simplest forecasting model and can be remarkably good for short-term horizons [17]. This model is the most

common reference model for short forecasting term of solar irradiance [18].

4.1. Analysis of a Case on 22nd June 2014. Now, we present an example of the simulated results on 22nd June 2014 from 12:52 until 12:56 using the new algorithm. The deviation

TABLE 2: Statistic indicator comparison between the new ANN forecast model and the persistence forecast model for the 50 manually selected pictures.

Simulation's time	Model	RMSE (cloud amount)	MAE (cloud amount)	R^2
5 minutes (from 145 different pictures)	ANN	51	43	0.96
	Persistence	70	61	0.79

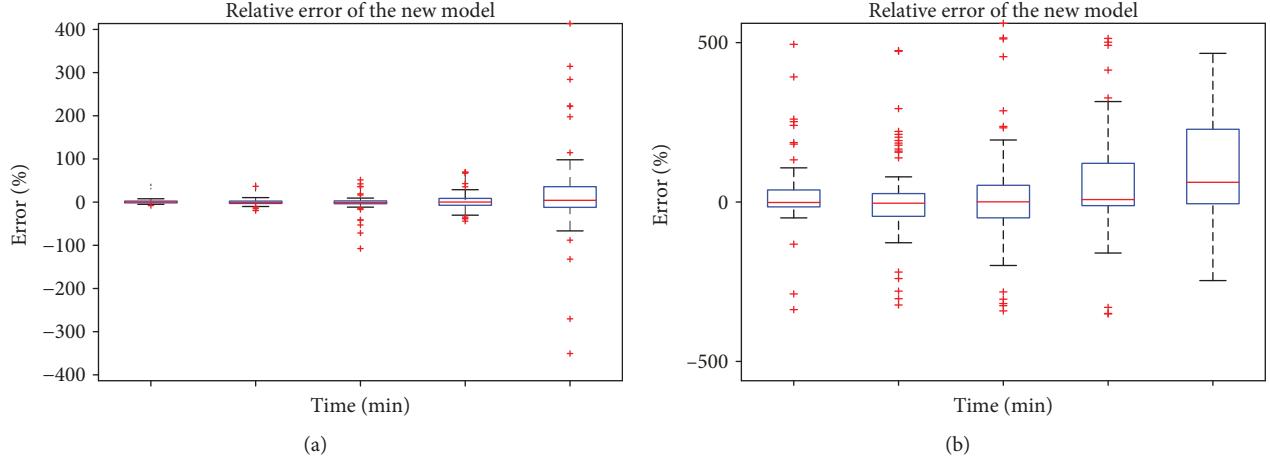


FIGURE 7: Relative deviation as boxplot for the first five minutes. (a) corresponds to the new ANN. Here, we can see that the deviations are narrower concentrated in the middle interquartile ranges. (b) corresponds to the persistence model. 50% of the deviations are not exactly located in the middle. In addition, the 25% and 75% of the deviation is higher than in (a).

between the measured cloud concentration at time $t + 1$ and the simulated cloud concentration at time $t + 1$ was calculated according to equations 6 and 7. The new algorithm produced a high-quality forecast compared to the reference persistence model.

The measured cloud concentration at 12:52 was 84.2%, and our algorithm simulated 75.8%, i.e., a difference of 8.4%, while the persistence model had a difference of 11.7%. For the next minute simulation (at 12:53), our model simulated 79.2% of the total cloud concentration and the measured value was 70.0% resulting in a deviation of 9.2%. Here, the persistence model difference is 13.9%. Figure 5 shows the measured cloud concentration of the first five minutes and the simulated cloud concentration for the new algorithm and the persistence model.

It is also worth mentioning that forecasting with images completely covered with gray and dark gray clouds is of minor relevance for solar energy forecast. Hence, images with a solar global irradiance smaller than 100 W/m^2 were not used in this work. Figure 6 shows the total deviation of both models.

Table 2 presents a comparison between the results of the different methods. Over 5 months of validation periods, we got 240 valid cases. However, not every picture was considered for validation. Full cloudy pictures (stratus cloud) and when the clouds did not have a form (shape) to be followed were not considered. Thus, only pictures with defined clouds (cumulus cloud) were considered for validation. To validate our algorithm, we applied statistical samplings. Therefore,

taking into consideration a confidence level of 95%, with a margin of error of 6%, our simulated cases were 145.

Therefore, applying the new ANN model to the 145 pictures, the presented model achieved an average of 30% for all sky conditions compared with the persistence model. Unfortunately, direct comparisons with other methods are difficult due to different time horizons and regional weather conditions.

Figure 7 shows the relative deviations as boxplots. The results suggest that the new model (Figure 7(a)) shows a symmetrical approach for the 50% sample rate. In addition, Figure 7(b) shows an asymmetrical distribution of outliers and a decreasing number of outliers, which leads to higher uncertainties. In conclusion, the uncertainty of the new model increased, but not as abruptly as with the persistence model, when more simulated data are introduced.

5. Conclusions

A new algorithm to forecast cloud distribution five minutes in advance has been presented. The model presented here combines all-sky images and an Ar-BP ANN. The cloud pixels were identified with the help of the Haze Index. The methodology described here only needs one all-sky image for predicting cloud concentration one minute ahead. According to the simulation results, our model makes a significant progress to predict cloud concentration five minutes in advance using a machine learning method, outperforming the persistence model. This method has already been

successfully tested as a tool as an important step for predicting the GHI one hour in advance [12]. The horizon time prediction of the new forecasting model can play an important role in German markets and within the European Union as well.

Future work will expand this methodology for forecasting the full image for longer periods, maybe using satellite information. In addition, the idea is to extend the proposed methodology to collect universally high-quality data giving a more robust validation.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

We would specially like to thank the German Academic Exchange Service (DAAD) for financing this investigation. The publication of this article was funded by the Open Access fund of the Leibniz Universität Hannover.

References

- [1] A. T. Lorenzo, W. F. Holmgren, and A. D. Cronin, "Irradiance forecasts based on an irradiance monitoring network, cloud motion, and spatial averaging," *Solar Energy*, vol. 122, pp. 1158–1169, 2015.
- [2] F. Wang, Z. Zhen, Z. Mi, H. Sun, S. Su, and G. Yang, "Solar irradiance feature extraction and support vector machines based weather status pattern recognition model for short-term photovoltaic power forecasting," *Energy and Buildings*, vol. 86, pp. 427–438, 2015.
- [3] https://www.bundesnetzagentur.de/DE/Sachgebiete/ElektrizitaetundGas/Unternehmen_Institutionen/Versorgungssicherheit/Engpassmanagement/Regelenergie/regelenergie-node.html.
- [4] V. Quaschning, "Statistiken," 2017, <https://www.volker-quaschning.de/datserv/pv-welt/index.php>.
- [5] IEA, *International Energy Agency, World Energy Outlook 2016, Executive Summary*, International Energy Agency, World Energy Outlook, 2016, http://www.iea.org/publications/free_publications/publication/WEB_WorldEnergyOutlook2015ExecutiveSummaryEnglishFinal.pdf.
- [6] D. Bernecker, C. Riess, E. Angelopoulou, and J. Hornegger, "Continuous short-term irradiance forecasts using sky images," *Solar Energy*, vol. 110, pp. 303–315, 2014.
- [7] R. Marquez and C. F. M. Coimbra, "Intra-hour DNI forecasting based on cloud tracking image analysis," *Solar Energy*, vol. 91, pp. 327–336, 2013.
- [8] J. Yang, W. Lu, Y. Ma, and W. Yao, "An automated cirrus cloud detection method for a ground-based cloud image," *Journal of Atmospheric and Oceanic Technology*, vol. 29, no. 4, pp. 527–537, 2012.
- [9] A. Cazorla, F. J. Olmo, and L. Alados-Arboledas, "Development of a sky imager for cloud cover assessment," *Journal of the Optical Society of America A*, vol. 25, no. 1, pp. 29–39, 2008.
- [10] S. Liu, L. Zhang, Z. Zhang, C. Wang, and B. Xiao, "Automatic cloud detection for all-sky images using superpixel segmentation," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 2, pp. 354–358, 2015.
- [11] E. Scolari, F. Sossan, and M. Paolone, "Irradiance prediction intervals for PV stochastic generation in microgrid applications," *Solar Energy*, vol. 139, pp. 116–129, 2016.
- [12] C. Crisosto, M. Hofmann, R. Mubarak, and G. Seckmeyer, "One-hour prediction of the global solar irradiance from all-sky images using artificial neural networks," *Energies*, vol. 11, no. 11, article 2906, 2018.
- [13] J. O. Kamadinata, T. Lit Ken, and T. Suwa, "Global solar radiation prediction methodology using artificial neural networks for photovoltaic power generation systems," in *Proceedings of the 6th International Conference on Smart Cities and Green ICT Systems - Volume 1: SMARTGREENS*, Porto, Portugal, 2017.
- [14] Z. Zhen, Z. Wang, F. Wang, Z. Mi, and K. Li, "Research on a cloud image forecasting approach for solar power forecasting," *Energy Procedia*, vol. 142, pp. 362–368, 2017.
- [15] M. Yamashita, M. Yoshimura, and T. Nakashizuka, "Cloud cover estimation using multitemporal hemisphere imageries," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science*, vol. 35, pp. 826–829, 2004.
- [16] M. Schrempf, *Entwicklung eines Algorithmus zur Wolkenerkennung in Digitalbildern des Himmels*, Diploma Thesis at IMuK, 2012.
- [17] D. S. Wilks, *Statistical Methods in the Atmospheric Sciences*, Academic Press, Cambridge, MA, USA, 2011.
- [18] M. P. Almeida, O. Perpiñán, and L. Narvarte, "PV power forecast using a nonparametric PV model," *Solar Energy*, vol. 115, pp. 354–368, 2015.

