Made available by Hasselt University Library in https://documentserver.uhasselt.be

Sky Images for Short-Term Solar Irradiance Forecast: A Comparative Study of Linear Machine Learning Models Peer-reviewed author version

Shirazi, Elham; GORDON, Ivan; Reinders, Angele & Catthoor, Francky (2024) Sky Images for Short-Term Solar Irradiance Forecast: A Comparative Study of Linear Machine Learning Models. In: IEEE Journal of Photovoltaics, 14 (4), p. 691 -698.

DOI: 10.1109/JPHOTOV.2024.3398365 Handle: http://hdl.handle.net/1942/43307 Sky Images for Short-Term Solar Irradiance Forecast: A Comparative Study of Linear Machine Learning Models E. Shirazi, *Member, IEEE*, I. Gordon, A. Reinders, *Senior Member, IEEE*, F. Catthoor,

1 Abstract-In this study, sky images for short-term solar irradiance forecasting are evaluated by means of seven linear machine learning algorithms. Namely an accurate solar irradiance forecast is critical to the reliable operation of power systems with the increasing integration of PV systems. In the first step, several features are extracted from sky images and reconstructed, and next used as exogenous inputs to seven machine learning algorithms, i.e. linear regression, ridge regression, Lasso regression, Bayesian ridge regression, stochastic gradient descent, generalized linear model regression and RANSAC. A representative dataset of three years of sky images with one minute resolution from 2014 to 2016, serves for comparison together with the clear sky indexes as inputs to forecast ground-level solar radiances for up to 30 minutes ahead. The results of the abovementioned algorithms are compared, where for 5 and 10 minutes ahead, Lasso has the highest accuracy with RMSE of 0.05 and 0.062 kW/m², while for 15 to 30 minutes ahead, stochastic gradient descent provides the most accurate forecast with RMSE of 0.067, 0.071, 0.074 and 0.076 kW/m² for 15,20,25 and 30 minutes ahead horizons respectively. For all time horizons, Bayesian ridge is among the three most accurate models and RANSAC has the highest error. The results show that ground level solar irradiance can be forecasted with a relatively low average instantaneous error ranging from 0.05 to 0.1 kW/m² depending on the model and forecasting horizon without imposing a too high execution time overhead namely less than 7 seconds. The accuracy of the forecast can be improved if combined with cloud detection algorithms. Overall ridge, Bayesian ridge and stochastic gradient descent provide more accurate forecasts for short-term horizons.

Index Terms— Solar forecast, sky imager, machine learning, short-term forecast.

E. Shirazi was with Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium, and Imec, imo-imomec, Thor Park 8320, 3600 Genk, Belgium. She is now with University of Twente, 7522 NB Enschede, The Netherlands (e-mail: e.shirazi@utwente.nl). I. Gordon is with Hasselt University, imo-imomec, Martelarenlaan 42, 3500 Hasselt, Belgium, and Imec. imo-imomec, Thor Park 8320, 3600 Genk, Belgium. (ivan.gordon@imec.be). A. Reinders is with Eindoven University of Technology and Solliance, P.O.Box 513, 5600 MB Eindhoven, The Netherlands. (a.h.m.e.reinders@tue.nl).

F. CATTHOOR IS WITH KATHOLIEKE UNIVERSITEIT LEUVEN, KASTEELPARK ARENBERG 10, 3001 LEUVEN, BELGIUM, AND IMEC, IMO-IMOMEC, THOR PARK 8320, 3600 GENK, BELGIUM. (CATTHOOR@IMEC.BE). I. INTRODUCTION

A mong all factors that influence PV power production, solar irradiance has the highest impact, which is influenced by geographic location, sun position, atmospheric conditions and the cloud thickness and height. Specifically, on the short run the dynamic changes in cloud shape and motion lead to fast fluctuations in irradiance and subsequently in PV power.. Hence, in order to accurately forecast irradiance in a short-term horizon, it is important to analyse clouds, for which high-resolution images of the sky are needed. A sky imager consists of either an imager looking downward onto a hemispherical mirror or a digital camera with fisheye lens which takes high resolution pictures of the sky providing the opportunity to track fast changes of solar irradiance in a short term. **Figure 1** shows three sky images captured by the VIVOTEK sky imager installed at the rooftop of the EnergyVille campus in Belgium [1].



Figure 1. Sky images captured under different weather condition at KU Leuven/Imec, EnergyVille Campus [1]

Short-term solar irradiance and power forecasts have become popular research topics over past decade. In this section we briefly discuss studies done on these topics using sky imagers. For instance, a sky imager-based solar irradiance forecast by a multi-step algorithm including cloud detection and motion, using two months of sky images data is proposed in [2]. In addition to sky images, seven cloud classes including stratocumulus, cirrocumulus, cumulus. altocumulus, nimbostratus, cumulonimbus, stratus, altostratus, cirrostratus, cirrus, and clear sky, are used as inputs to the forecasting algorithm. The cloud classification has been done using 600 manually classified images and applying support vector classification (SVC). The data from 99 pyranometer is used for validation of the proposed method which at the end did



overall process of the proposed met

All of the abovementioned studies tackled the forecasting problem; however, they have either used complex methods (e.g., CNN-MLP) or time consuming approaches (e.g., manually labelling sky images) to conduct forecasting or the forecasting horizons were limited (e.g., 180 seconds). Therefore, the aim and first main contribution of this study is to show that we can still achieve accurate forecast without very complicated methods, using carefully selected inputs to meet the forecast application needs. To meet this requirement, not only spectral features but also textural features have been extracted from sky images, reconstructed, and then used as exogenous input to the forecasting model in addition to clear sky indexes. The forecasting has been done for ground-level solar irradiance rather than a specific PV system output, so the result could be used by many different users in the region, from power system operators to private PV system owners. Moreover, the forecasts have been done for a range of horizons from 5 to 30 minutes ahead rather than a specific horizon using seven different machine learning algorithms. The second main contribution is that we carefully compare these seven methods for our context and discuss the obtained insights including their advantages and weak points.

This study is structured as follows: Section II will introduce the proposed method for analysing irradiance data together with sky imager-based forecast leading to answering research questions of short-term solar irradiance forecasting. The results are shown in section III. The paper will be completed with a discussion in section IV and conclusion in section V.

II. PROPOSED METHODOLOGY

The overall process of the proposed method is demonstrated in Figure 2, which consists of two parallel processes. The first process starts with capturing sky images and correcting their fish-eye lens distortion (Figure 2, Block #1). In the next step, specific features are extracted from these sky images. The value of the Red (R), Green (G) and Blue (B) colour channels, together with the red-to-blue ratio and the normalized red-toblue ratio are the five extracted features from the sky images (Figure 2, Block #2). In the next steps, three vectors of entropy, average -, and standard deviation are calculated for each of these features, so in total there will be 15 vectors as inputs to the forecasting model from the first process (Figure 2, Block #3). The second process is to calculate the clear sky index for GHI and Direct Normal Irradiance (DNI), using the GHI and DNI time series together with the clear sky model (Figure 2, Block #4). In this study, the Ineichen and Perez clear sky model has been used including Linke turbidity to represent the transparency level of the atmosphere [6]. After calculating the clear sky irradiance using the Ineichen and Perez model, the clear sky index is calculated by dividing the measured irradiance at ground level to the clear sky estimated irradiance. The variability, lagged average, and backward average are calculated for the series of clear sky indexes (Figure 2, Block #5). These three vectors in addition to the 15 vectors from the first process are used as inputs for the machine learning algorithms (Figure 2, Block #6). In the following, sky image feature reconstruction (A) and machine learning algorithms (B) will be described.

A. Sky images feature reconstruction

In this step, specific sky image features are extracted and then reconstructed to be used in the forecast model. These features could be either spectral or textural and are quantified characteristics such as image textures and colour values. These features help to discriminate clear sky from clouds which will have a significant influence on solar irradiance on the ground level.

Colour is the main property used in sky imagers-based forecasts. Spectral features are based on colour models and are more stable than contour and edges as they are global features [7]. The solar irradiance forecasting models are mainly based on blue and red channels, as clouds mostly scatter red and blue, whereas clear sky scatters higher red values. The spectral features considered in this study are the red, blue, and green channels, the normalized red to blue ratio and the red to blue ratio. For each of these features, the mean and standard deviation are calculated by (1) and (2).

(1) (2)

where is the colour channel, is the value of each channel in pixel and is the number of total pixels in a sky image.

The texture of a sky image can be calculated from Grey Level Co-occurrence Matrices (GLCM) which is a square matrix based on the histogram of the image [8]. In this study entropy has been included in the model which is a measure of randomness of grey level differences and computed from the histogram of each colour channel separately [9].

(3)

Here is the element on row a and column b of GLCM. It represents the relative frequency that two pixels occur.

B. Machine Learning Algorithm Options

There are many factors accounting for determining the ground level solar irradiance, some of which can be determined through explicit equations like the angle of incident and solar zenith angle, while the others are constantly changing and are difficult to be calculated through physical models, such as clouds. As it has been discussed, dynamic changes in clouds form and cloud motion result in fluctuations of solar irradiance especially on the short-term. Machine learning algorithms can help to perceive such complex association without any explicit equations. That is one of the reasons which lead to increasing application of machine learning based forecasting in recent years.

The machine learning models can be either linear or nonlinear. The former assumes there is a linear relationship between inputs and outputs, while the latter assumes a nonlinear one. Linear models use convex optimization approaches and are simple, fast, and straightforward, while non-linear models are more complex and slower. In this study multiple Linear Machine Learning Algorithms (LMLA) have been considered (Figure 2, Block #6). The aim is to investigate the effect of independent variables , namely sky images and clear sky indexes, on the average value of dependant variable that is solar irradiance. The general model of LMLA is as follows, where the expected value of is a function of :

2

(4)

(5)

By decomposing (4) the dependant variable can be modelled as follows, where is the random deviation from the expected values.

Linear regression [10], ridge regression [11], least absolute shrinkage and selection operator (Lasso) regression [10], Bayesian ridge (BR) regression [11], stochastic gradient descent (SGD) [12], generalized linear model (GLM) [13] regression and random sample consensus (RANSAC) [14] are among LMLA which have been applied on the dataset to forecast solar irradiance. The differences between these models lay in the objective functions and the added penalty terms. In the linear regression, the objective function is to minimize the sum of squared errors between the forecasted (and target values in the dataset with no penalty term, while in the Lasso and ridge regressions penalty terms of the L1 and L2 regularizations are added to the objective function respectively [10]. L1 is the sum of absolute values of regression coefficients, while L2 is the sum of squared regression coefficients [11]. Bayesian ridge also considers L2 as the penalty term, but with a randomly assigned variable which controls the amount of shrinkage. The gradient descent method has the same objective function as the linear regression, on the other hand it optimizes the regression coefficients through an iterative method [12]. The normal gradient descent tends to become slow with large datasets and fall into a local minimum, stochastic gradient descent [15] on the other hand prevents these from happening by introducing a noisy gradient of a single or a minibatch of data points [16]. Like ridge and Bayesian ridge regression, generalized linear model regression uses L2 as the penalty factor as well, assuming has a form of an exponential distribution [13]. The last model is Random Sample Consensus (RANSAC) regressor, which is different from other models. It defines a hypothesis using randomly chosen data points and evaluates the hypothesis using the remaining data points in the dataset [14]. If the number of data points satisfying the defined hypothesis is less than a specified threshold, then a new hypothesis has to be defined. This iteration continues until the total number of satisfying data points exceed a threshold, or a maximum number of iterations is reached [17].

In addition to abovementioned models, other models have also been considered such as Huber Regressor [18], Theil-Sen Regression [19] and Decision Tree Regressor [20], however they either did not converge or resulted in high errors; therefore, they were deemed not suitable for our case study.

III. RESULTS

In order to evaluate the performance of the proposed method, a dataset containing three years of sky images and irradiance measurements has been used which is not recorded in EnergyVille campus, but in Folsom city [21]. The irradiance data are measured using a second-generation Rotating Shadow band Radiometer (RSR) sensor from Augustyn, Inc with one minute resolution. The sky camera captures RGB colour image with 1536x1536 pixels at one minute intervals. The dataset is divided into two subsets, namely training dataset (2/3) and test dataset (1/3). The first

two years are used for training the models and the data of the last year is used for testing the models.

The LMLA have been applied using the Scikit package in Python [22]. A smart persistence algorithm has been used as the benchmark for this comparison. The persistence method assumes the value of irradiance in the next step is equal to the value of irradiance at the current step, while in the smart persistence method, a scaling factor is added as follows:

(6)

can be either GHI or DNI at time and is the scaling factor. is the backward average of the ratio of measured irradiance to clear sky irradiance of previous steps.

(7)

The results of short-term forecasting for 5 minutes ahead of GHI and DNI have been presented in Figure 3 and Figure 4. The other horizons including 10, 15, 20, 25 and 30 minutes ahead also follows the same pattern but with higher errors.





As can be seen, all algorithms can follow the measurements, however the RANSAC algorithm has many ripples. There are two potential explanations, the first is that RANSAC converts the forecasting problem into a selection problem. The second possible explanation is that RANSAC is sensitive to hyperparameters tuning and the parameters are not optimized here. Regarding those reasons, it is concluded that RANSAC cannot provide an accurate short-term forecast with sky imagers under these circumstances unless it is optimized. It can be observed that DNI has higher error in comparison to the GHI, which is aligned with existing literature [23]. This is mainly due to the effect of aerosol optical depth (AOD) on the irradiance components.

A. Forecast Accuracy

In order to check the accuracy of each model, the Root Mean Square Error (RMSE) has been considered as the main metric. RMSE is a common metric to measure the accuracy of a forecasting model. It is defined as follows, where and are the forecasted and measured values respectively.

(8)

The GHI and DNI forecasting error for each method over all horizons from 5 to 30 minutes ahead are shown in Figure 5



Figure 5. RMSE of GHI and DNI forecasting via different models

As expected, the RANSAC algorithm does not perform well, while the stochastic gradient descent, ridge and lasso regression have the best performance. It should be noted that the smart persistence algorithm uses the updated value for clear sky irradiance instead of the value of the last step, so it can perform rather accurate especially in shorter-term horizons i.e., 5 minutes. Another factor which should be considered while comparing error is the variance of the error. In case of all models except for RANSAC, the variance is lower than smart persistence.

Three most accurate models based on RMSE for each horizon, have been selected. For 5 and 10 minutes ahead, Lasso has the highest accuracy, while for 15 to 30 minutes ahead, stochastic gradient descent provides the most accurate forecast. In all horizons, Bayesian ridge is among the top three most accurate forecasts. The effect of the forecasting horizon on the accuracy of the forecast is shown in Figure 6.



Figure 6. GHI and DNI forecasting error for different horizons The limited field of view of the sky imagers limits the forecasting horizons of such methods. Normally sky imager based methods do not provide accurate forecast for horizons of more than 30 minutes, depending on cloud and wind speed.

This is derived based on the previous studies of the same group. [5], [24]. An increasing forecasting horizon clearly increases the uncertainty and hence the errors.

Three years of data from 2014 to 2016 have been used in this study, where the data from 2014 to 2015 are used as training dataset and the data for 2016 used as the test dataset. The average RMSE for the training dataset is 0,068 and for test dataset is 0,070 [kW/m2].



Figure 7. GHI and DNI forecasting error for test and train datasets As demonstrated, the errors of both test and training datasets are quite close, which means that the proposed algorithm could accurately fit the data and there is no over-fitting or under-fitting problem. It is also observed that the error of the test dataset is lower, which proves that there is no bias in diving the dataset into two subsets of train and test.

A comparison of the results of this study with existing literature is insightful. In order to do so, three comparable studies have been chosen. In [25], a CNN is applied to forecast 5-20 min ahead of GHI using sky images and lagged GHI. The results show RMSE of 49-177 W/m², 93-146 W/m², 71-118 W/m² in sunny day, partly cloudy day and overcast day, respectively. [26] presents the results of 2.5 and 5-minutes solar forecasting system based on two sky-imagers in Canary Islands, Spain, by identifying clouds and predicting their movement. The results of two days namely a very cloudy day with high variability in irradiance, and a sunny day. The first day has a RMSE of 939.8 W/m2, while the second day had a lower RMSE of 240.9 W/m2. In [27], a short-term DNI forecasting method using sky imagers is proposed. It considers both the cloud coverage and the influence of other atmospheric particulates such as the absorbing, reflecting, and scattering of DNI. The results of DNI forecasting for 3 to 7 minutes ahead for 6 different summer days in June, July, and August show different RMSEs ranging from 103 W/m2 to 355 W/m2. Comparing these results with the results of our study, shows that the RMSE of different models are relatively low.

Although RMSE is a useful metric to evaluate the performance of the models, it cannot sufficiently describe the performance. Another index is model bias which refers to the systematic error that results from certain assumptions in the modelling process. Unlike random errors, which can vary unpredictably, bias describes a consistent or persistent distortion from the measurement. Comparison with baseline models helps to identify potential biases by setting a benchmark for performance, highlighting systematic errors, validating improvements, and indicating data-related issues.

This comparison ensures that the complexities of the forecasting models are indeed contributing to better, more unbiased predictions rather than merely fitting to noise or reinforcing existing prejudices in the data. Different methods can be used to address model bias, such as data preprocessing, feature engineering and model tuning. In this study, feature engineering is used to extract and create the most relevant features that better capture the underlying patterns in the data (Figure 2). In addition to that a baseline model is used to compare all models' performances which will be discussed in the next section.

C. Forecast Skill

Another indicator which can be used for evaluating the performance of forecasting is the Forecast Skill (FS), where the persistence model is used as the reference model.

(9)

Here the smart persistence model is used as the reference, which has higher accuracy in comparison to the persistence method. All forecasting models perform better than the smart persistence algorithm for short term horizons from 5 to 30 minutes, except for the RANSAC model. Considering the causes which have been explained earlier, it is concluded that RANSAC cannot provide an accurate short-term forecast with sky imagers under these circumstances, further investigation regarding optimized RANSAC performance in forecasting solar irradiance is needed.

To have a better comparison, the RANSAC method has been removed from the comparison, as the values for forecasting skill are mostly negative. The forecasting skills of other models have been compared in Figure 11 and Figure 12. As it is demonstrated there, GLM has lower skills in comparison to other methods, besides RANSAC. It might be because in GLM methods the predictor variables should be uncorrelated, which in our case they are not. There are also other possible explanations, such as GLM sensitivity to outliers, strict assumptions around distribution shape or its dependency on the unknown parameters of the fitted model which is one of the limitations of GLM. Various techniques such as the quantile dispersion graphs of the mean-squared error can be used to deal with this problem.



Figure 8. GHI and DNI forecasting skills for different models As it is demonstrated, Lasso can provide more accurate forecasts in total, with comparatively same forecasting skill, followed by Bayesian ridge and ridge regression. The

difference between Lasso and ridge regression is in the penalty terms. Lasso imposes penalty using linear proportional values of the errors while ridge uses squared values. Both of these methods have lower variance and higher accuracy in comparison to linear regression and smart persistence for 5 to 30 minutes ahead.

From 15 minutes ahead onward, SGD is amongst the most accurate methods. It could be due to its ability to set learning rate of the algorithm as a function of iteration number. In this way the algorithm makes significant changes in the beginning and fine tune the parameters in the later iterations.

D. Ramp Rate Detection

Ramp rate in solar forecasting refers to the rate at which solar power output changes over a specific time. It is a critical parameter for grid operators because sudden increases (rampup events) or decreases (ramp-down events) in solar power generation can significantly affect grid stability and energy management strategies. In solar forecasting, models aim to predict these ramp rates accurately to inform grid operators about potential rapid changes in solar generation. This enables them to take proactive measures, such as managing demand response strategies, to ensure grid reliability and balance supply and demand effectively. In order to evaluate the ability of the models to capture changes in irradiance and therefore solar PV power, the RR is calculated based on the measurements and forecasts using equation (10).

(10)

Where is the irradiance at time t [W/m2], and is the time between two consecutive steps in minutes. If RR is above or below certain thresholds, then a ramp-up or ramp-down event is detected and therefore an appropriate control measure such as charge/discharge storage or curtailment should be taken. As shown in Figure 9 the models can capture these changes with relatively low RMSE. As the forecasting horizon increases, the ramps become smoother and therefore the RMSE decreases.



Figure 9. RMSE of RR for different forecasting horizons

As shown in Figure 10 and Figure 11 among the models which are explored in this study, GLM regression has the best ability to capture ramp rates and therefore ramp events with the lowest RMSE namely 11.20 and 11.21 for 30 and 25 minutes ahead forecasts of GHI respectively. The RANSAC

model on the other hand is the worst when it comes to forecasting ramp rates with RMSE of 71.59 and 102.73 for 25 and 30 minutes ahead forecasts of DNI respectively.



Figure 10. RMSE of RR for different forecasting models



Figure 11. RMSE of RR for different forecasting models

E. Execution Time

One of the reasons for choosing linear machine learning models is the execution time. These algorithms are simpler and do not impose high execution time.



Figure 12. Average execution time of different models

AS DEMONSTRATED IN FIGURE 9, THE LASSO MODEL HAS THE HIGHEST EXECUTION TIME, WHILE IT IS AMONG THE MOST ACCURATE MODELS. THE MOST SUITABLE MODEL CAN BE

CHOSEN FOR THE APPLICATION CONSIDERING EXECUTION TIME AND ACCURACY OF THE MODEL. IV. DISCUSSION

Although, the proposed models showed overall higher accuracy in comparison to its competitor for short-term forecast, i.e., the persistence model, it should be noted that various weather conditions can impact the accuracy of the forecast. For example, there are more fluctuations and consequently higher errors in cloudy weather in comparison to sunny weather. One idea is to train a model specifically for each weather condition and use that model for the forecast under that specific weather condition.

The sky imager based forecast can provide high spatial resolution if the sky images have high quality. Low-quality sky images can present additional bias. Thus, calibrating sky imager parameters such as exposure time, focal length, and geometric alignment are quite important prerequisites. All abovementioned factors can influence the quality of the image and therefore the quality of the training data, which plays an important role in the performance of machine learning based forecasts. In addition to the quality of the data, the quantity of the data is also important. To train a machine learning algorithm effectively, at least two years and to test the algorithm at least one year of data is required [28], which means in total at least three years of data on a specific site is required to have a reliable accurate forecast. This is one restricting factor with such methods. That is the main reason why we have selected the Oldenburg data set in this paper, because it meets these requirements. Sky imagers provide information on the dynamics of clouds which has the biggest influence on the irradiance. To improve the accuracy of forecast, cloud detection algorithms and integrated cloud tracking techniques can be also added to the forecasting model [7]. Although sky imagers have high spatial and temporal resolution, their spatial coverages are limited, and location dependant. Exploring the possibility to adopt the forecast model of a location in another location is one possible direction for future research. It would also be advisable to extend the number of locations with different climate and more sky imagers in order to increase the validity of the results.

In this study, the Ineichen and Perez clear sky model [6] has been used to calculate the clear sky index, however there are various clear sky models such as ASHRAE, BIRD, Heliosat, DPP, MAC, Kasten, REST2, MAGIC (SARAH-2), and SOLIS [29], [30]. These models can include various factors in their models, which influence the clear sky index and therefore the inputs to the forecasting model. Conducting a sensitivity analysis on different clear sky models can show the robustness of the forecasting algorithm to different clear sky indexes.

V. CONCLUSION

In order to have a reliable operation of power systems with high penetration of PV, it is essential to have an accurate solar irradiance forecast. This research developed a solar irradiance forecasting method based on machine learning algorithms. For that purpose, sky imager data for three years are used as exogenous input to multiple machine learning algorithms to forecast global horizontal and direct normal irradiances for 5 [21]

to 30 minutes ahead. The results of the seven different models have been compared and the three most accurate models have been retained for short-term horizon namely ridge, Bayesian ridge and stochastic gradient descent. Bayesian ridge, stochastic gradient descent and generalized linear model regression are the fastest algorithms in our case.

The results show that ground-level solar irradiance can_1be forecasted with a relatively low error ranging from 0.05 to 0.1 kW/m² using linear machine learning models without imposing a too high execution time overhead, namely less than 7 seconds. There is a trade-off between execution the and accuracy of the algorithm which will influence the choice of the most suitable algorithm for a specific application. Nevertheless, both Bayesian ridge regression and stochastic gradient descent are among the fastest and most accurate models.

REFERENCES

- [1] "Research into sustainable energy and intelligent energy systems | EnergyVille." [Online]. Available: https://www.energyville.be/en. [Accessed: 29-Mar-2022].
- T. Schmidt, J. Kalisch, E. Lorenz, and D. Heinemann, "Evaluatin^[22] [2] the spatio-temporal performance of sky-imager-based solar irradiance analysis and forecasts," Atmos Chem Phys, vol. 16, no. 5, pp. 3399-3412, Mar. 2016.
- [3] M. Lipperheide, J. L. Bosch, and J. Kleissl, "Embedded nowcasting method using cloud speed persistence for a photovoltaic power [23] plant," Solar Energy, vol. 112, pp. 232-238, Feb. 2015.
- [4] O. el Alani, M. Abraim, H. Ghennioui, A. Ghennioui, I. Ikenbi, and F. E. Dahr, "Short term solar irradiance forecasting using sky images based on a hybrid CNN-MLP model," Energy Reports, vol. 7, pp^[24] 888-900, Nov. 2021.
- [5] D. Anagnostos, T. Schmidt, S. Cavadias, D. Soudris, J. Poortmans, and F. Catthoor, "A method for detailed, short-term energy yield forecasting of photovoltaic installations," Renew Energy, vol. 130, [25] pp. 122-129, Jan. 2019.
- [6] P. Ineichen and R. Perez, "A new airmass independent formulation for the Linke turbidity coefficient," Solar Energy, vol. 73, no. 3, pp. 151-157, Sep. 2002.
- [7] E. Shirazi, J. Lemmens, M. Gofran Chowdhury, A. Tuomiranta, F. Catthoor, E. Voroshazi, and I. Gordon, "Cloud Detection for PV [26] Power Forecast based on Colour Components of Sky Images," Conference Record of the IEEE Photovoltaic Specialists Conference, pp. 2389–2391, Jun. 2021.
- R. M. Haralick, I. Dinstein, and K. Shanmugam, "Textural Features [8] for Image Classification," IEEE Trans Syst Man Cybern, vol. SMC-3, no. 6, pp. 610-621, 1973.
- [9] A. Heinle, A. Macke, and A. Srivastav, "Atmospheric Measurement?" Techniques Automatic cloud classification of whole sky images." Atmos. Meas. Tech, vol. 3, pp. 557-567, 2010.
- L. Kahane, "Regression Basics," *Regression Basics*, Dec. 2014. A. E. Hoerl and R. W. Kennard, "Ridge Regression: Biased [10]
- [28] [11] Estimation for Nonorthogonal Problems," Technometrics, vol. 42, no. 1, p. 80, Feb. 2000.
- T. Zhang, "Solving large scale linear prediction problems using ^[29] [12] stochastic gradient descent algorithms," Proceedings, Twenty-First International Conference on Machine Learning, ICML 2004, pp. [30] 919-926, 2004.
- P. McCullagh and J. A. Nelder, Generalized linear models. Boca [13] Raton ; Chapman and Hall, 1989.
- [14] V. K. Saini and A. Maity, "Random sample consensus in decentralized Kalman filter," Eur J Control, vol. 65, p. 100617, May 2022.

L. Fahrmeir, T. Kneib, S. Lang, and B. D. Marx, Regression, Models, Methods and Applications. Berlin, Heidelberg: Springer Berlin Heidelberg, 2021.

W. Lizhen, Z. Yifan, W. Gang, and H. Xiaohong, "A novel shortterm load forecasting method based on mini-batch stochastic gradient descent regression model," Electric Power Systems Research, vol. 211, p. 108226, Oct. 2022.

M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," Commun ACM, vol. 24, no. 6, pp. 381-395, Jun. 1981.

E. M. Ronchetti and P. J. Huber, "Robust statistics, 2nd edition," Wiley, 2009.

P. K. Sen, "Estimates of the Regression Coefficient Based on Kendall's Tau," J Am Stat Assoc, vol. 63, no. 324, pp. 1379–1389, 1968.

L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, Classification and regression trees, CRC Press, 2017.

H. T. C. Pedro, D. P. Larson, and C. F. M. Coimbra, "A comprehensive dataset for the accelerated development and solar forecasting methods," Journal of benchmarking of Renewable and Sustainable Energy, vol. 11, no. 3, p. 036102, Jun. 2019.

F. Pedregosa, V. Michel, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, J. Vanderplas, D. Cournapeau, F. Pedregosa, G. Varoquaux, A. Gramfort, B. Thirion, O. Grisel, V. Dubourg, A. Passos, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

C. A. Gueymard, "Temporal variability in direct and global irradiance at various time scales as affected by aerosols," Solar Energy, vol. 86, no. 12, pp. 3544–3553, Dec. 2012.

I. K. Bazionis, M. A. Kousounadis-Knousen, P. S. Georgilakis, E. Shirazi, D. Soudris, and F. Catthoor, "A taxonomy of short-term solar power forecasting: Classifications focused on climatic conditions and input data," IET Renewable Power Generation, vol. 17, no. 9, pp. 2411-2432, Jul. 2023.

A. Ryu, M. Ito, H. Ishii, and Y. Hayashi, "Preliminary Analysis of Short-term Solar Irradiance Forecasting by using Total-sky Imager and Convolutional Neural Network," 2019 IEEE PES GTD Grand International Conference and Exposition Asia, GTD Asia 2019, pp. 627-631. May 2019.

D. Canadillas, B. Gonzalez-Diaz, J. Rodriguez, J. Rodriguez, and R. Guerrero-Lemus, "A low-cost two-camera sky-imager ground-based intra-hour solar forecasting system with cloud base height estimation capabilities working in a smart grid," 2018 IEEE 7th World Conference on Photovoltaic Energy Conversion, WCPEC 2018 - A Joint Conference of 45th IEEE PVSC, 28th PVSEC and 34th EU

PVSEC, pp. 2282-2287, Nov. 2018. H. Zhang, C. Zhang, and W. Li, "Short-Term DNI forecasting using

cloud history patterns matching," 7th International Conference on Information Science and Technology, ICIST 2017 - Proceedings, pp. 57-63, May 2017.

C. Feng, J. Zhang, W. Zhang, and B. M. Hodge, "Convolutional neural networks for intra-hour solar forecasting based on sky image sequences," Appl Energy, vol. 310, p. 118438, Mar. 2022.

M. J. Reno, C. W. Hansen, and J. S. Stein, "Global Horizontal Irradiance Clear Sky Models: Implementation and Analysis, SAND2012-2389," Albuquerque, New Mexico 87185, 2012.

F. Antonanzas-Torres, R. Urraca, J. Polo, O. Perpiñán-Lamigueiro, and R. Escobar, "Clear sky solar irradiance models: A review of seventy models," Renewable and Sustainable Energy Reviews, vol. 107, pp. 374-387, Jun. 2019.