Renewable Energy 187 (2022) 1009–1019

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

A network of sky imagers for spatial solar irradiance assessment

Yinghao Chu [a, b, c], Mengying Li [a, d], Hugo T.C. Pedro [a], Carlos F.M. Coimbra [a, *]

[a] Department of Mechanical and Aerospace Engineering, University of California San Diego, La Jolla, CA, USA
[b] College of Electronics and Information Engineering, Shenzhen University, Guangdong, China
[c] Accenture Labs, Shenzhen, Guangdong, 518000, China
[d] Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hong Kong, China

ARTICLE INFO

Article history:
Received 23 May 2021
Received in revised form
29 December 2021
Accepted 10 January 2022
Available online 3 February 2022

Keywords:
Sky-imaging network
Image-to-irradiance
Solar resourcing
Spatially-resolved irradiance

ABSTRACT

A network of seven low-cost hemispheric sky-imaging cameras has been installed in the Los Angeles basin. This network of cameras provides wide sky coverage to perform spatial solar irradiance assessments. An Image to Irradiance algorithm (I2I) is proposed to simultaneously derive high-resolution diffuse, direct and global solar irradiance from sky images. Spatial interpolation using the Kriging method is used to derive the irradiance field for the whole basin area. The relatively inexpensive network of cameras can provide spatially resolved GHI that is more accurate than GHI derived from GOES-west satellite images provided by the Cooperative Institute for Meteorological Satellite Studies (CIMSS) when the distance to the nearest site is less than 40 km. This work successfully demonstrates that, with minor trade-off in accuracy, solar irradiance monitoring can be achieved using off-the-shelf cameras in the absence of radiometers.

© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

California utilities are experiencing unprecedented growth of distributed solar energy in order to align with the goal for utilities to generate 50% of power from renewable sources by 2030 [1,2]. The dramatic growth in Photovoltaic (PV) installations will introduce significant variability to the electric grid due to the variable nature of solar irradiance [3]. If short-term (particularly intra-hour horizons) solar irradiance can not be accurately assessed, the costs of the integration of solar power will remain costly because of the excess investment on energy storage or regulation capacity, particularly for micro-grids and distribution feeders [4,5]. Major components of ground level solar irradiance include Global Horizontal Irradiance (GHI), Direct Normal Irradiance (DNI), and Diffuse Horizontal Irradiance (DHI). High resolution GHI assessment is particularly useful for the design and operation of PhotoVoltaic (PV) system while DNI assessment is of great importance for Concentrated Solar Power (CSP) systems.

To assess short-term high-resolution solar irradiance, different radiometers such as pyranometers or pyrheliometers have been developed based on either thermopiles or photodiodes sensors [6,7]. To assess GHI, DNI, DHI all together using only one device, Rotating Shadowband Irradiometers (RSIs) have been proposed [8]. However, the rotating shadowband or shadowball will decrease the sampling frequency of irradiance measurements, and precise controls of the shadowband to track and block the Sun will significantly increase the maintenance costs of deployed RSIs.

Assessing solar irradiance using sky imaging systems is a potential approach to address the above issues. Sky imaging systems play important roles in aerosol characterization, cloud detection, and solar forecasting [9]. Sky imagers provide spatial-temporal information on the dynamics of clouds, which is the dominating factor affecting ground-level irradiance. Multipurpose sky imagers are also used for calculation of cloud base height [10], automatic clear sky detection [11], distribution analysis of atmospheric scatters [12], and cloud cover assessment for meteorological applications [13]. If all components of solar irradiance are assessed by cost-effective sky imagers in place of additional solar instrumentation, the overall cost of telemetry can be noticeably decreased. Such systems have been proposed in the literature [9,14,15]. However, available studies either depend on expensive total sky imagers or are unable to simultaneously estimate all components of solar irradiance. Recently, researchers have used inexpensive imaging devices to measure irradiance and other atmospheric variables. For instance, Kocifaj et al. [16] used all-sky camera images to derive upward emission and Garstang's emission function. Hänel et al. [17]

* Corresponding author.
E-mail address: ccoimbra@ucsd.edu (C.F.M. Coimbra).

https://doi.org/10.1016/j.renene.2022.01.032
0960-1481/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

proposed the use of calibrated consumer digital cameras with fisheye lenses to measure night sky brightness. For day-time solar irradiance measurement, Scolari et al. [18] estimated GHI using principal component analysis and artificial neural network based on images obtained from an all-sky camera. Herrera et al. [19] proposed a GHI estimation scheme based on a linear regression using RGB sky image features. Jiang et al. [20] used a CNN based image regression approach to calculate solar irradiance during cloudy periods. Valentín et al. [21] developed an irradiance nowcast method based on low-cost sky imager. In addition to nowcast, Kamadinata et al. [22] used an ANN to predict GHI for 1–5 min in advance from sky images.

In many real world applications, irradiance assessment for a single point in space is not sufficient. For instance, operators of large centralized solar installations need to understand the solar resource over the solar farm, which usually covers areas of several kilometers square. Utility operators need to assess the distributed solar generations over large areas (e.g. cities) to estimate the amount of underlying demand and the penetration of distributed solar power [23]. Consequently, spatial assessment of solar irradiance is essential for planning, integrating, regulating, and managing solar power generations. However, few studies develop image-based irradiance measurement methods for large areas.

To address these issues, an Image to Irradiance algorithm (I2I) is developed in this work to derive all components of solar irradiance (GHI, DNI, and DHI) from images captured by low-cost fish-eye cameras. This fish-eye camera (named as SkyCam in this work) is equipped with a CMOS sensor, which captures sky images in which the pixel color intensity is highly correlated to the solar irradiance values [24,25]. Based on the I2I algorithm, a network of SkyCams is developed for high-resolution intra-hour spatial solar assessment. The network of SkyCams is deployed to assess minute-wise solar irradiance field (via the I2I model) for the Los Angeles basin using kriging spatial interpolation technique, whose performance is validated by cross validation and compared against satellite-derived GHI (temporal resolution of 15-min). The proposed SkyCam network system is useful to develop accurate low-cost spatial solar assessments for both large centralized solar plants and distributed solar power systems over a large area.

The Image to Irradiance methodology is introduced in Section 2. The derivation of irradiance field over the LA basin area is presented in Section 3. The validation of the proposed I2I algorithm is presented in Section 4. Conclusions are followed in Section 5.

2. Image to irradiance model for SkyCam

In this section, an Image to Irradiance (I2I) model is developed to derive solar irradiance values from sky images. In this model, relevant image features are extracted from SkyCam images, and the diffuse horizontal irradiance (DHI), direct normal irradiance (DNI) and global horizontal irradiance (GHI) are derived based on these image features.

2.1. Data preparation and feature engineering

For a plane normal to the zenith, both direct and diffuse irradiance components are received from the hemispherical sky. Therefore in this work, a low-cost Vivotek model FE8173V camera (denoted as “SkyCam”) is used to capture sky images with 180°Panoramic View and 360°Surround View. The FE8173V is a fish-eye fixed dome network camera featuring a 3.1 MP-resolution sensor. It is weatherproof and suitable for monitoring outdoor areas. The specifications of the camera are presented in Table 1. The SkyCam captures a sky image every minute and transfers it via FTP through 4G data network to a server located at University of California San Diego. The captured images are 1536 by 1536 RGB-based JPG images that can be directly interpreted by the proposed algorithm. The pixel intensities of RGB channels range from 0 to 255. This process is fully automated, requiring only occasional cleaning of the camera dome. In addition to regular dome cleaning, about 5% of archived images are randomly selected for manual inspections. Images with dirt or drops are removed from the model development phase to ensure good image quality.

The I2I model is developed at a single independent (San Diego, California) location and then applied to the network of cameras. The model uses 1-min sampled sky images captured with a SkyCam and DHI, GHI and DNI co-located measurements acquired a Multi-Filter Rotating Shadowband Radiometers (MFR-7, Yankee Environmental Systems). Images are captured using two different settings for the camera: from July 20 to August 20, 2015, the SkyCam was set to a constant exposure time of 1/32 000 s (the lowest setting of this camera); from August 20 to September 20, 2015, it was set for an automated exposure time. Nighttime images and irradiance measurements are discarded. Training and testing datasets were separated based on the week number, with data from odd weeks assigned to the training set and the remainder assigned to the testing set. Both the training and testing sets include diverse weather and cloud-cover conditions.

The I2I model considers, for each time instance, the current image and the five previous 1-min sampled images, thus considering cloud motion and cloud fraction information. In searching for informative image features, five masks are applied to each image, as shown in Fig. 1. Fig. 1 (a) shows an example of a sky image, (b) shows the projected image (to eliminate fish-eye distortion [16,26]) and (c) shows cloud velocity vectors derived by Particle Image Velocimetry (PIV) method [27]. The next five panels show the aforementioned five masks:

- a strip mask in the reverse direction of average cloud motion – Fig. 1(d);
- a grid element from the strip mask in which the distance from the grid element to the Sun is the cloud speed times the time lag – Fig. 1(e);
- a disk mask whose boundary is perpendicular to the cloud velocity (it includes possible clouds that are moving towards the Sun) – Fig. 1(f);
- a mask that only contains the Sun and circumsolar region, where Sun location is found by a Sun locating algorithm proposed in Ref. [25] – Fig. 1(g); On the projected images, the radius of the mask is 25 pixels, which is about 5° half-angle of view. The current recommended half-angle of the World Meteorological Organization (WMO) is 2.5° [28]. The half-angle of the mask is larger than the WMO recommendation in order to account for any error in the sun locating algorithm.
- and a mask that contains only the sky dome, excluding circumsolar region and surrounding obstacles – Fig. 1(h);

The features used to model irradiance are extracted from seven color representations for each image [29,30]. They are: red channel (R), green channel (G), blue channel (B), hue (H), saturation (S), value (V) and intensity (I). All are scaled to be in range [0, 1]. For each of the masks presented above, 30 image features (Table 2) are calculated. For each time instance, the total number of numerical variables obtained from sky images is (number of features per mask) × (number of masks) × (number of lagged images) = 30 × 5 × 6 = 900.

Table 1
SkyCam specifications.

Weight	0.6 kg	**Dimensions**	145 mm (D), 52 mm(H)
Temp. Range	−25 to 50 °C	**Power**	3.84 W peak
Voltage	12 VDC	**Connector**	Power-over-Ethernet or standard 3-prong US AC plug

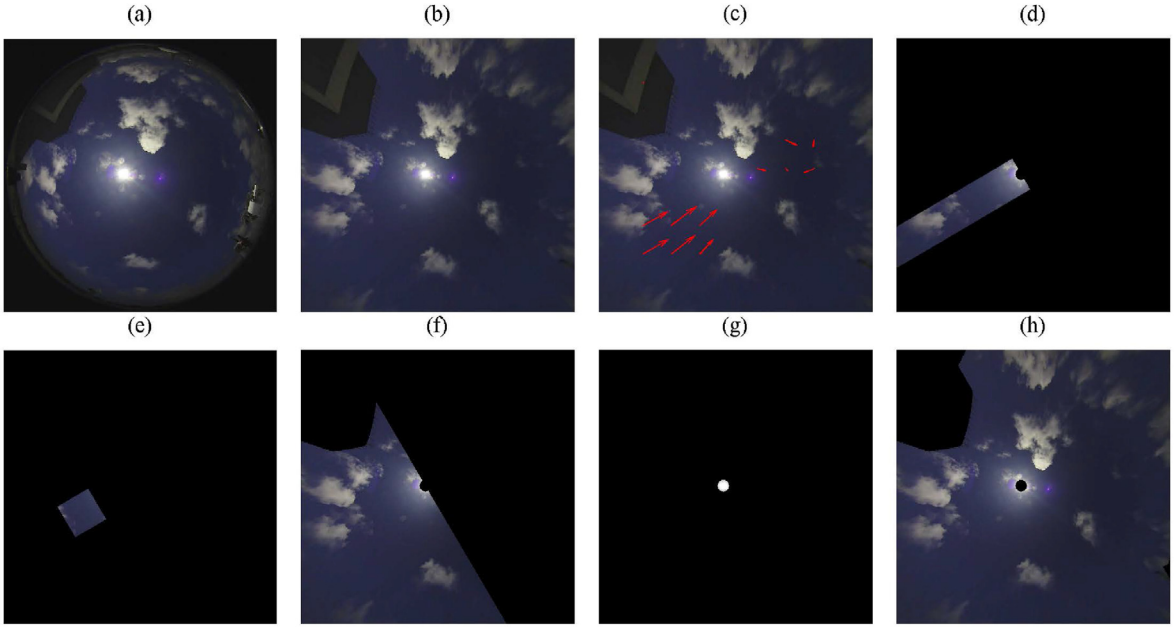

Fig. 1. Image pre-processing before the derivation of irradiance. (a) original image, (b) projected image, (c) cloud moving velocity, (d)ladder mask, (e) grid mask, (f) disk mask, (g) sun mask and (h) sky mask.

2.2. Derivation of diffuse horizontal irradiance

Given that DHI results from the light scattered in the atmosphere, it is reasonable to consider only features obtained from the mask that include all the sky dome (Fig. 1(h)). Therefore, DHI modeling relies on $30 \times 6 = 180$ numerical variables at each time instance. Relationship between the measured DHI with the best correlated features for dynamic exposure images and constant exposure images is shown in Fig. 2. As shown in Fig. 2, the image features obtained with constant exposure have a much stronger linear correlation with DHI than features from dynamically exposed images. Therefore, the default exposure setting (which is dynamic) of the SkyCam should be changed to have a constant exposure time for I2I applications.

Given the high linear correlation between DHI and some of the image features, we explore a simple linear model to derive 5-min averaged DHI from the 180 variables:

$$\hat{I}_d = a_0 + \sum_{i=1}^{180} a_i F_i \tag{1}$$

where a_i are regressed coefficients and F_i are image features.

To account for nonlinearity, a MultiLayer Perceptron (MLP) model is also explored. The MLP is a popular stochastic learning tool for pattern recognition, data classification and regression, and has proven to be useful for non-linear input/output mapping [31]. The weights and bias of MLP are estimated using the training data set by a supervised learning process [32,33]. In this work, the MLP places the basic processing elements (neurons) in layers and allows only forward connections of the neurons [34]. Layers between the first input layer and the last output layer are hidden layers. Neurons take in weighted sum of inputs X_j from various layers, add bias, and produce an output using an activation function,

$$Y_i = f\left(\sum_{j=1}^{M}(w_{ij}X_j + \beta_{ij})\right), \tag{2}$$

where Y_i is the output from the $i-$ th neuron, and $f(\cdot)$ is the activation function. The structure of the MLP follows the recommendation of previous works [30,35]: the number of hidden layer is set to 1 and the number of neurons per hidden layer is set to 7. Once the learning process is finished, the MLP model generates derived DHI using new inputs (selected image features).

For a proper comparison between features obtained with fixed and dynamic exposure settings, the model is applied to both cases. Several error metrics (defined in Appendix A) for the DHI model

Table 2
Image features calculated in each mask.

Features	Descriptions
1–7	mean of R,G,B,H,S,V,I
8–14	standard deviation of R,G,B,H,S,V,I
15–21	entropy of R,G,B,H,S,V,I
22	mean of (R−B)/(R + B)
23	mean of (R−B)
24	mean of (R•I)
25	mean of 1/3(R + G + B)
26	mean of (R−B)•S
27	mean of (R−B)•H
28	mean of H•S
29	mean of R•V
30	mean of B•V

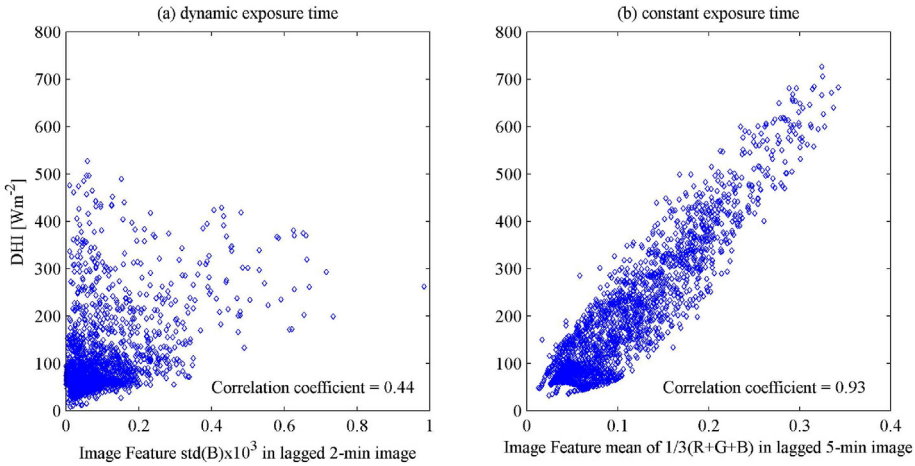

Fig. 2. DHI with respect to best correlated image feature for (a) cameras with default dynamic exposure time and (b) cameras with constant exposure time. By setting the exposure time to be a constant, the image feature is much better correlated with the irradiance values as indicated by the values of correlation coefficients.

based on the two data sets are presented in Table 3. As expected from Fig. 2, the errors are much smaller for features from constant exposure settings. The values listed in the table demonstrate that the linear model, despite its simplicity, performs marginally better than the MLP model. Thus, in the remainder of this paper, we adopt the linear model using features from constant exposure images to estimate DHI.

Fig. 3 plots the derived DHI and measured DHI time series and the respective instantaneous error. As shown in the figure, the derived DHI effectively captures the measured values. It is worth mentioning that this model also performs well in periods of large variability in which DHI ramps up or down suddenly.

2.3. Derivation of global horizontal irradiance

Since GHI is the combination of both the direct and diffuse irradiance components, features from the five masks that include both the sky dome and the circumsolar region are considered in the I2I model leading to 900 predictors at each time instance. Similar to DHI derivation, a linear model and an MLP model are explored to derive GHI clear sky index. The GHI clear sky index k_t is defined as GHI divided by clear sky GHI [35,36], where clear sky GHI is obtained using Ineichen's clear sky irradiance model [37]. The lowest exposure time setting is chosen to minimize the pixel saturation issues, especially in the circumsolar area. The aforementioned methods may suffer from relatively large errors when the Sun is obscured by thin clouds, during which the circumsolar pixels are still saturated but the ground-level GHI or DNI decrease significantly. This situation will be further investigated in our future works.

Fig. 4 plots the correlation coefficient between GHI's clear sky index and the 900 image features. The absolute values of the correlation coefficients are plotted in Fig. 4 in the form of a color map. The labels in the x-axis indicate the mask that was applied and the values in the y-axis indicate the lag of the image/feature. The color map indicates that sun mask features have the highest correlation coefficients with GHI clear sky index, followed by sky mask. Features derived with the other three masks (disk mask, strip mask and ladder mask) rarely show large correlations, and are, therefore, not included in the set of predictors in the I2I model.

Table 4 presents the results for GHI derivation as a function of the model (linear or MLP) and 12 sets of predictors described in the table. The results indicate that, in this case, the MLP model outperforms the linear model and that using only features obtained with the Sun mask and sky mask (Model 7) leads to the lowest GHI estimation error. The RMSE of Model 7 is 3.2% less than that of the second best Model 6, and is 10.1% less than that of the median performing model. Therefore Model 7 was selected for further analysis in this work. For data captured at different conditions or by different devices, Model 7 may not be the best one. Therefore, we provided details of other potential models in Table 4 that may better suit other scenarios. The I2I GHI time series sample is plotted in Fig. 5.

2.4. Derivation of direct normal irradiance

The methodology to derive DNI is similar to that of GHI derivation. Once clouds are present, the influence of aerosols becomes less significant. Therefore, we choose to use a clear-sky model that has built-in AOD information to normalize DNI. Then our I2I is focused on inferring cloud optical depth (COD) from the image features. Trial and error experiments with the training data set leads to an MLP model that uses 360 image features from the sun and sky masks to derive DNI clear sky index. A sample of the derived DNI time series is plotted in Fig. 6, indicating that the model effectively captures the ramps of DNI as well. Note that DNI

Table 3
I2I performance of DHI derivation.

	Dynamic Exposure	Contant Exposure	Dynamic Exposure	Contant Exposure
	Linear model	Linear model	MLP model	MLP model
MBE [W m^{-2}]	−11.1	2.6	−1.8	−5.7
MAE [W m^{-2}]	46.1	29.0	40.6	31.8
RMSE [W m^{-2}]	73.2	39.2	69.5	44.9
rRMSE [%]	58.2	16.4	55.3	18.8

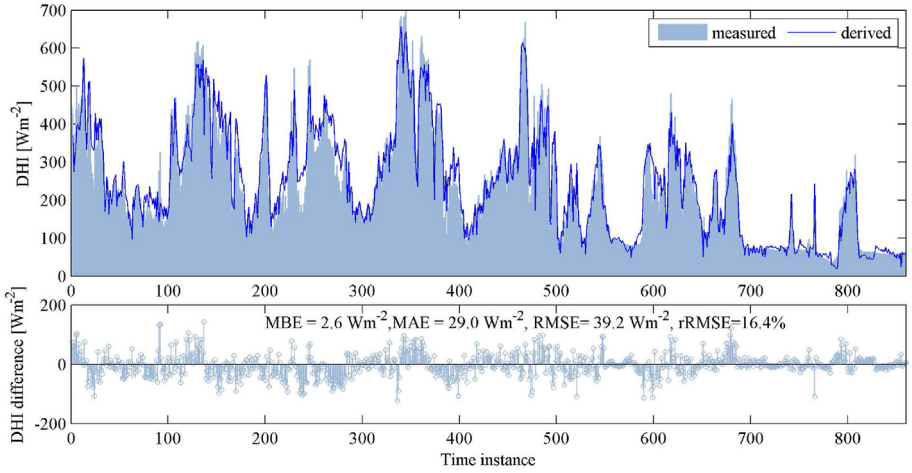

Fig. 3. Sample derived DHI time series compared against measured DHI for a few days in the testing set.

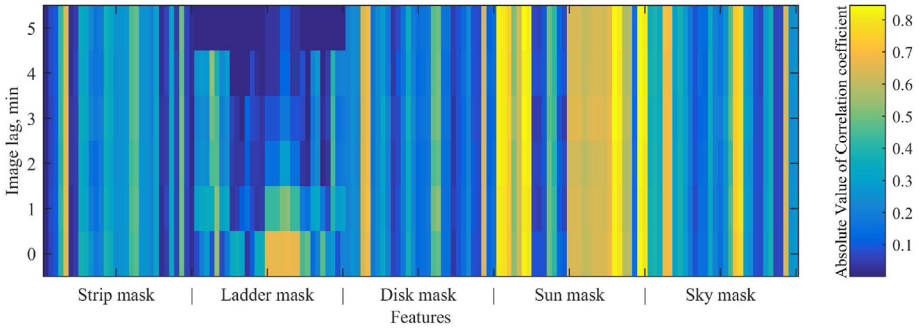

Fig. 4. Absolute value of the correlation coefficients of image feature and GHI clear sky index, for the 5 masks in 6 lagged images.

Table 4
Input selection for GHI derivation.

error matrix	Linear	MLP model											
		1*	2*	3*	4*	5*	6*	7*	8*	9*	10*	11*	12*
MBE [W m^{-2}]	1.0	19.4	−2.1	14.5	2.5	−8.7	6.4	**−8.4**	4.8	−2.7	3.0	−0.9	8.3
MAE [W m^{-2}]	64.7	66.2	71.7	76.8	62.5	62.5	53.0	**51.6**	57.2	56.0	56.2	57.2	56.8
RMSE [W m^{-2}]	86.4	95.8	96.1	109.4	89.1	85.8	75.6	**73.2**	78.9	77.8	78.1	78.0	80.6
rRMSE [%]	15.4	14.2	16.7	16.1	15.9	15.1	13.6	**13.0**	14.2	13.8	14.0	13.9	14.6

1* 900 inputs: all features.
2* 180 inputs: strip mask features.
3* 180 inputs: ladder mask features.
4* 180 inputs: disk mask features.
5* 180 inputs: Sun mask features.
6* 180 inputs: sky mask features.
7* 360 inputs: Sun and sky masks features.
8* Sun and sky masks features, when correlation coefficient > 0.7.
9* Sun and sky masks features, when correlation coefficient > 0.6.
10* Sun and sky masks features, when correlation coefficient > 0.5.
11* 300 inputs: Sun and sky masks, 5 lagged images (0 to 4-min lagged).
12* 420 inputs: Sun and sky masks, 7 lagged images (0 to 6-min lagged).

usually has larger variability than GHI, as such, the error of the I2I DNI model (listed in the error plot in Fig. 6) is larger than that of GHI.

As observed in Figs. 3, 5 and 6, the relatively large errors are usually associated with more dynamic weather when irradiance time series show significant fluctuations. Therefore, the error not only arises from the I2I algorithm, but also from potential synchronization mismatch between pyranometer and sky image, particularly when the irradiance fluctuates violently. A properly calibrated sensor may still encounter similar errors. In future work, we will consider the application of calibrated cameras to enhance the performance.

3. Assessing spatial GHI over an area using a network of cameras

Sky images for spatial GHI derivation are captured using SkyCams deployed in the Los Angeles basin from April 1 to November 1,

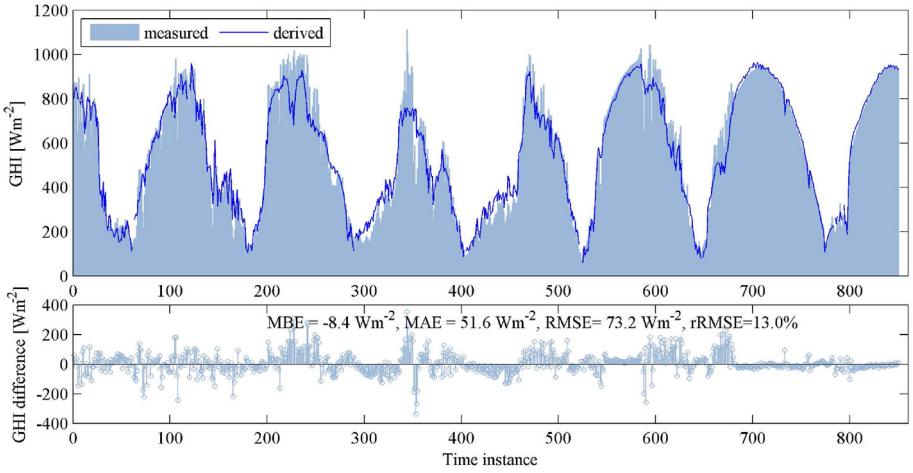

Fig. 5. Sample derived GHI time series. The I2I GHI was obtained with the MLP model with the 7* set inputs as listed in Table 4.

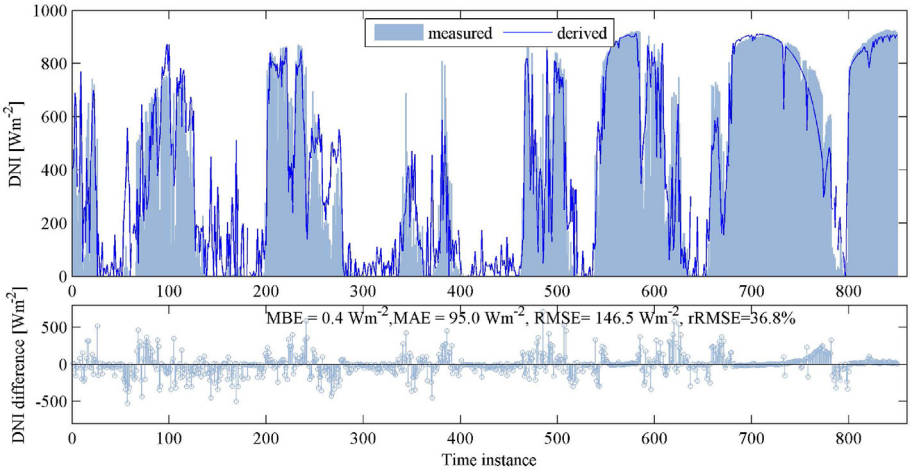

Fig. 6. Sample derived DNI time series. The I2I DNI was obtained with the MLP model with the 7* set inputs as listed in Table 4.

2017. The exposure time of all the SkyCams has been set as constant, according to the discussion above. Also, in the results presented below, nighttime measurements have been discarded.

3.1. Monitoring sites and data

The monitoring sites for the SkyCam network were chosen based on feasibility, site accessibility, and good field of view (that is, no major obstructions from trees or buildings). Table 5 presents the information of the monitoring sites.

3.2. Spatial interpolation methods

Ideally high-resolution grid solar irradiance could be obtained with a dense network of sensors. In practice, this is unattainable due to high cost, maintenance requirements, and adverse terrain conditions [38]. Alternatively, interpolation techniques based the measurement from a few distributed monitoring sites can be used to estimate irradiance field in a large geographical domain [39,40].

Kriging interpolation [41] is one of the geostatistics methodologies for valuating, estimating, and studying spatial characteristics of a regionalized variable [42]. Also known as the best linear unbiased estimator [43], kriging estimates the value Z at arbitrary sites as the weighted sum of Z_i at known sites i. The mathematical expression of the representative ordinary kriging is [42]:

$$\hat{Z}_0 = \sum_{i=1}^{N} w_i Z_i, \tag{3}$$

where N is the number of known sites, w_i is the unbiased weights:

$$\sum_{i=1}^{N} w_i = 1. \tag{4}$$

w_i will be derived using variogram model to minimize the variance of the estimation:

$$\sigma_0^2 = \mathrm{Var}[\hat{Z}_0 - Z_0], \tag{5}$$

under the constraint:

$$\mathrm{E}[\hat{Z}_0 - Z_0] = 0. \tag{6}$$

For two sites with separation distance d, the mean and the variogram of the difference in Z are assumed to be functions of d:

$$m(d) = \mathrm{E}[Z(x+d) - Z(x)] \tag{7}$$

Table 5
Monitoring sites of the SkyCams.

ID	Sites	Latitude, °	Longitude,°	Altitude, m	Images Available from	Apogee GHI data available from
a	Azusa	34.136	−117.924	187	September 9, 2015	N/A
b	Banning	33.921	−116.858	671	September 24, 2015	N/A
c	Fontana	34.100	−117.492	363	August 25, 2015	N/A
d	Glendora	34.144	−117.850	278	September 9, 2015	April 1, 2017
e	Pico Rivera	34.010	−118.069	58	September 17, 2015	April 1, 2017
f	Rubidoux	33.999	−117.416	248	September 9, 2015	April 1, 2017
g	Santa Clarita	34.383	−118.528	386	September 17, 2015	N/A

$$\gamma(d) = \mathrm{Var}[Z(x+d) - Z(x)], \tag{8}$$

where the variogram γ can be estimated using the known samples:

$$\gamma(d) = \frac{1}{N(d)} \sum_{i=1}^{N(d)} (Z(x+d) - Z(x))^2, \tag{9}$$

where $N(d)$ is the number of sample pairs with a distance d. Theoretically, the increment of the γ is finite and will attain a sill. To minimize the variance:

$$\begin{aligned} \sigma_0^2 &= \mathrm{E}\left[(\hat{Z}_0 - Z_0)^2\right] \\ &= \sum_i \sum_j w_i w_j \mathrm{E}[(Z_i - Z_0)(Z_j - Z_0)] \\ &= \sum_i \sum_j w_i w_j \gamma(x_i - x_j) + 2\sum_i w_i \gamma(x_i - x_0), \end{aligned} \tag{10}$$

where x_i represent the $i-$ th known site, and the first-order derivation of σ_0^2 with respect to each w_i equals to zero corresponds to the minimum condition:

$$\sum_j w_j \gamma(x_i - x_j) + \lambda = \gamma(x_i - x_0), \tag{11}$$

which can be expressed in matrix form:

$$\boldsymbol{\Gamma w} = \boldsymbol{\gamma}_0, \tag{12}$$

where $\boldsymbol{\Gamma} = \begin{Bmatrix} \gamma_{11} & \gamma_{12} & \cdot & \gamma_{1N} & 1 \\ \gamma_{21} & \gamma_{22} & \cdot & \gamma_{2N} & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \gamma_{N1} & \gamma_{N2} & \cdot & \gamma_{NN} & 1 \\ 1 & 1 & \cdot & 1 & 0 \end{Bmatrix}$, $\gamma_{ij} = \gamma(x_i - x_j)$, $\boldsymbol{w} = [w_1, w_2, \ldots w_N, \lambda]^T$, and $\boldsymbol{\gamma}_0 = [\gamma(x_1 - x_0), \ldots \gamma(x_N - x_0), 1]^T$. The $\gamma_{ij}(i, j = 1, 2, \ldots, N)$ are the variograms between two known sites, and the $\gamma(x_i - x_0)$, $(i = 1, 2, \ldots, N)$ are the estimated variograms between a known site and an investigated site, which is calculated by Eq. (9). Afterward, the weights **w** can be calculated as:

$$\boldsymbol{w} = \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma}_0. \tag{13}$$

4. Model validation

4.1. Validation of derived GHI for individual sites

For the three sites (Glendora, Pico Rivera and Rubidoux) that have GHI sensors, a MLP model is trained for each site using data from odd numbered weeks. Then the three MLP models are assembled as an integrated I2I model to derive GHI for all other sites where GHI measurements are unavailable. Table 6 presents the GHI derivation error of the testing set (compared to measurements) of the integrated I2I model. The errors for the three sites are comparable to the errors presented in section 2.3, indicating that the performance of the I2I model is independent of location.

4.2. Validation of derived GHI field

In this case, the validation is performed following a leave-one-out approach. For each time instance in the testing set we compute the irradiance for one of the seven sites by interpolating the I2I data from the other six sites. The procedure is repeated for the other sites by leaving the respective data out of the kriging interpolation. This dataset is denoted as $\hat{I}_{krig}$.

To strengthen the analysis we also compare the data interpolated for the seven sites against GHI data derived from satellite imagery. For this purpose, we use data from the Cooperative Institute for Meteorological Satellite Studies (CIMSS). Using images from the GOES-West satellite, CIMSS creates cloud fields and irradiance maps (among other variables) using the CLAVR-x algorithm [44]. The CIMSS data is published in real time at the FTP repository ftp://ftp.ssec.wisc.edu/clavr/goes_west/processed/. From this repository it was possible to obtain CIMSS data for 890 time instances present in the testing set from April 23 to June 19, 2017. From these files we extracted GHI data for the seven sites which we denote as $\hat{I}_{CIMSS}$.

The accuracy of the whole test set $\hat{I}_{krig1}$ and the two data subsets $\hat{I}_{krig2}$ and $\hat{I}_{CIMSS}$ are assessed by comparing their values against the I2I data for the seven sites. The error metrics that result from this analysis are listed in Table 7. For the case of $\hat{I}_{krig}$, the table shows the values for the whole testing set and the values for the smaller set that coincides with CIMSS data. Note that no values are listed for site e for the smaller set because there were many time instances the SkyCam did not properly operate to provide image data. Table 7 also lists the distance from the site for which the interpolation is performed to the nearest site (column "Dist"). The values allow to directly compare GHI determined with kriging from I2I data against GHI derived from satellite imagery. In summary, the GHI interpolated with kriging from nearby sites is more accurate than satellite-derived data if the distance to the nearest site is not too large. This finding is in agreement with results reported by Perez et al. [45] that defined a ≈30 km threshold after which solar resource assessments based on satellite techniques are preferable to interpolation. The rRMSE variation with distance is more evident in Fig. 7.

Table 6
GHI derivation errors for three sites in the Los Angeles basin.

error matrix	Glendora	Pico Rivera	Rubidoux
MBE [Wm^{-2}]	12.4	−41.2	−10.2
MAE [Wm^{-2}]	41.1	57.6	38.8
RMSE [Wm^{-2}]	59.1	78.9	61.5
rRMSE [%]	11.2	15.3	11.2

Table 7
Error metrics for $\hat{I}_{krig}$ and $\hat{I}_{CIMSS}$ relative to I2I GHI for the seven sites.

Site	Dist	MBE			MAE			RMSE			rRMSE		
		Krig1[a]	Krig2[b]	CIMSS[b]	Krig1	Krig2	CIMSS	Krig1	Krig2	CIMSS	Krig1	Krig2	CIMSS
a	6.8	−5.3	−12.5	−57.6	28.6	42.6	101.8	52.4	70.8	124.1	9.9	11.6	20.4
b	52.2	41.3	50.1	−25.4	59.5	80.1	105.3	86.5	113.3	176.5	14.6	16.5	25.7
c	13.2	−5.0	−3.6	−24.4	38.7	49.9	75.1	67.6	82.0	101.0	12.1	13.1	16.1
d	6.8	−2.9	−3.2	−62.5	27.6	39.4	106.2	52.1	66.9	126.7	9.8	11.0	20.9
e	19.3	−17.9	–[c]	–	50.2	–	–	83.0	–	–	15.9	–	–
f	13.2	2.2	1.2	−11.9	43.5	56.8	80.9	77.1	92.8	110.6	14.0	14.6	17.4
g	59.2	46.1	51.7	31.6	60.6	73.3	69.1	84.8	107.6	99.5	14.3	15.7	14.5

[a] Values computed for the whole testing set $\hat{I}_{krig1}$.
[b] Values computed for the subset (890 instances) with $\hat{I}_{krig2}$ and $\hat{I}_{CIMSS}$.
[c] I2I GHI data not available for site e for this subset.

The figure also shows a curve fit for the $\hat{I}_{krig}$ rRMSE variation with distance ($R^2 = 0.85$).

Finally, Fig. 8 compares GHI maps for a few instances characterized by cloudy weather (top and middle panels) and a single instance characterized by few clouds (bottom panel). As mentioned, site e (indicated by a red symbol) is not used to produce the $\hat{I}_{krig}$ maps. The interpolation region around each site is limited to a 40 km radius. This value indicates the threshold after which satellite-derived data is more accurate than interpolated data. It should be based on a large dataset as done in Perez et al. [45]. Thus, the chosen value of 40 km, based on results show in Fig. 7, should be regarded as a preliminary estimate that may change with more sites and longer dataset.

The red brackets in the figure's colorbars (bottom of the panels) indicate the GHI range for the two data sets. As expected the $\hat{I}_{krig}$ shows a much smoother GHI map with lower variability and narrower bounds than $\hat{I}_{CIMSS}$. This indicates that, if GHI's spatial variability is more important for resource assessment, than average GHI values, the satellite-derived data may be preferable to the kriging GHI. Regardless, these results show that a relatively inexpensive network of cameras can provide spatially resolved data that is more accurate than data that results from the processing of satellite images, under the condition that the interpolation distance remains under ≈40 km. This distance threshold can be changed for different applications. For instance, Ruiz-Arias et al. [46] showed that ground stations distributed homogeneously with a spacing of 100–150 km are able to produce unbiased gridded estimates of *daily* GHI.

4.3. Limitations and application potential of the proposed methodology

Compared with available ground based pyranometers, the accuracy of the proposed method is improvable, particularly for application of scientific research. But on the other hand, this method has shown high potential in many real-world applications such as.

- Solar forecasting systems for distributed PV sites where forecasting hardware and installation costs must be minimized due to economic constraints.
- Building rooftops usually require complicated procedures to obtain hardware installation permits due to considerations of security, safety, water-proof, etc. Cameras are more likely to obtain the permit as a part of anti-theft systems when compared with pyranometers and other scientific devices.
- Remotely located solar power sites that need to measure a number of meteorological data (include but not limited to cloud base height, clear sky detection, aerosol, cloud cover, DNI, GHI) but with limited maintenance resources. To maintain data quality, pyranometers need periodic high standard calibration, while cameras only need occasional cleaning of the domes.

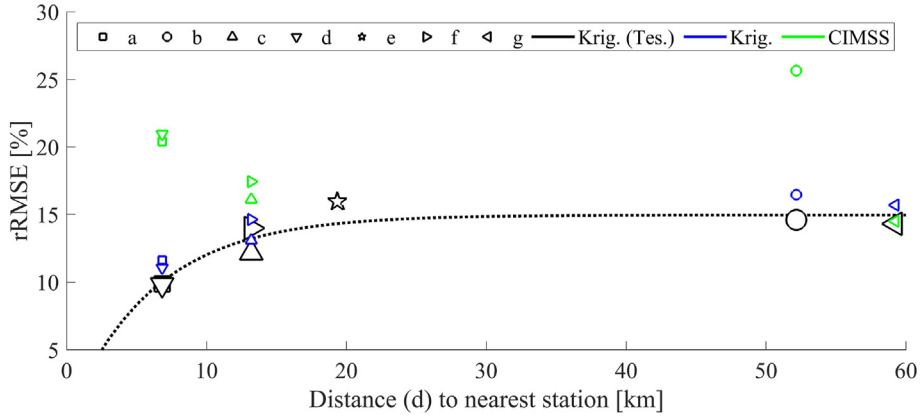

Fig. 7. rRMSE of $\hat{I}_{krig}$ (blue and black markers) and $\hat{I}_{CIMSS}$ (green markers) with respect to the I2I data for the seven sites. The black markers indicate rRMSE obtained for the whole testing set (14 212 data points), wheres blue and green markers indicate rRMSE for the subset in which $\hat{I}_{krig2}$ and $\hat{I}_{CIMSS}$ are available (890 data points). Different markers indicate the different sites. The x-axis indicates the distance to the nearest site. The dashed black line indicates a curve fit ($14.96[1 - \exp(-0.16\,d)]$, $R^2 = 0.85$) between the $\hat{I}_{krig1}$ rRMSE (for the whole testing data) and distance. For the CIMSS GHI data there is no evident trend with distance since it is derived from satellite images.

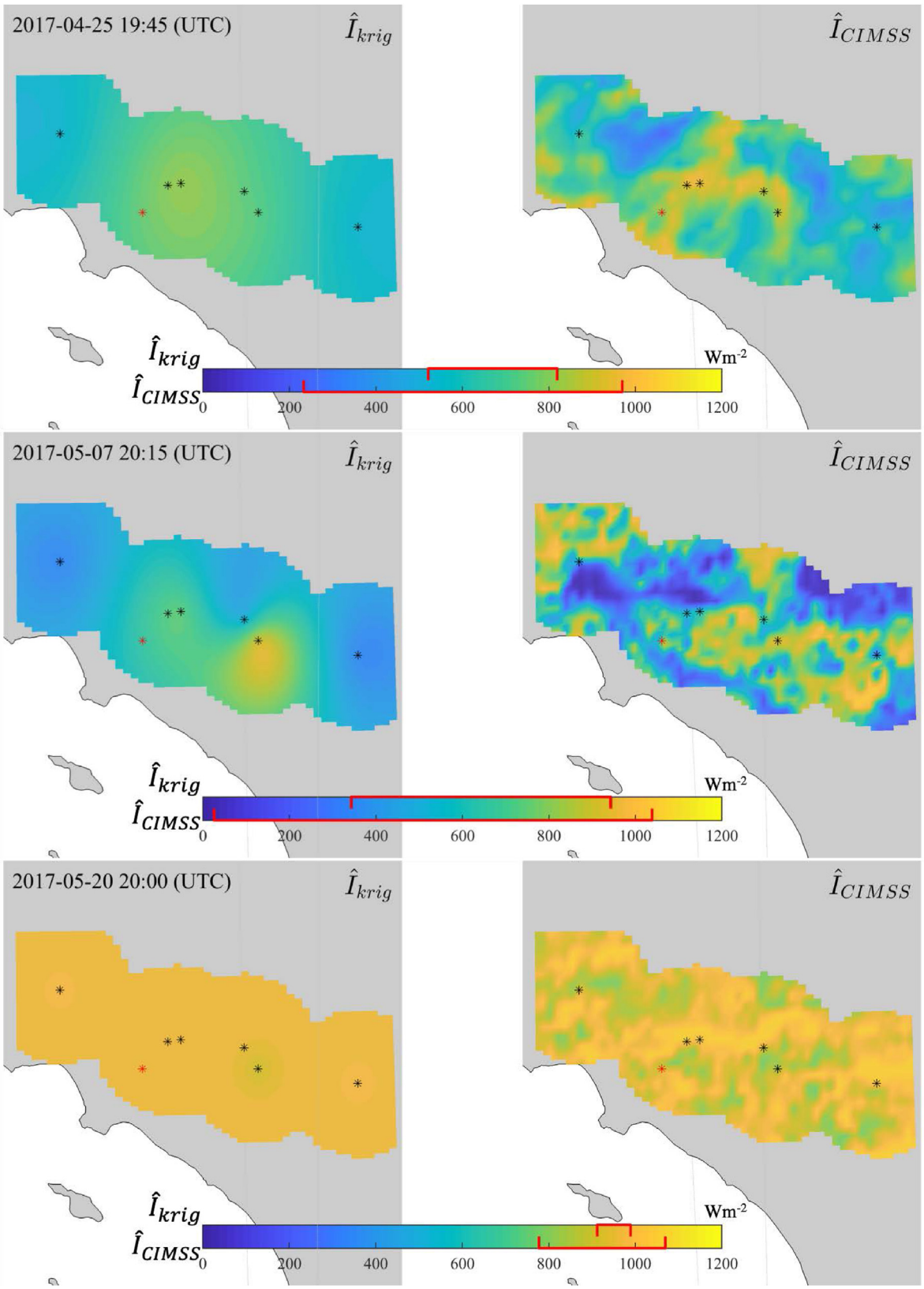

Fig. 8. GHI maps derived using kriging (left) and obtained from CIMSS (right) for three time instances close to noon (20:00 UTC). Maps are limited to a 40 km radius around available sites. The colorbar quantifies the GHI color map and the top (bottom) red brackets on the colorbar indicate the GHI range for $\hat{I}_{krig}$ ($\hat{I}_{CIMSS}$).

5. Conclusions

A sky imaging network that utilizes seven off-the-shelf fish-eye cameras (SkyCam) is deployed in the Los Angeles basin to evaluated the feasibility of using only imagers to produce intra-hour spatial solar irradiance assessment. An Image to Irradiance (I2I) model is proposed to derive ground-level diffuse horizontal irradiance (DHI), global horizontal irradiance (GHI) and direct normal irradiance (DNI) from the images with constant exposure time. With the proposed algorithm, off-the-shelf fish-eye cameras with proper settings can be applied as multi-functional and low-cost alternatives to radiometers for solar irradiance assessments applications with a little trade-off in measurement accuracy.

With the image-derived GHI from the seven sites in the Los Angeles basin, a kriging method is used to compute the spatial irradiance field over the basin. When compared with satellite-derived irradiance provided by CMISS using GOES-West satellite data, the kriging interpolated irradiance is generally more accurate then satellite-derived values when the distance to the nearest site is under 40 km. This proposed I2I model based imager-only network is potentially useful to develop short-term spatial solar irradiance assessment systems with substantially reduced instrumental and maintenance costs.

This work aims to develop a model that directly predicts irradiance from sky images through end-to-end data learning. In future work, we could use a cloud detection model to obtain a cloud map first and then use the unified model [47] tto calculate the irradiance to benchmark our proposed methods. Future work will also include but not limited to performance optimization through considering more physical constraints, model optimization and assessment with more data under greater diversity of weather conditions, and performance evaluation with irradiance derived from images of remote sensing based systems.

Data availability

The data used in this work is available upon reasonable request to the authors.

CRediT authorship contribution statement

Yinghao Chu: Formal analysis, Data curation, Writing – original draft, All authors contributed equally to this manuscript. All authors were involved in the planning of the research, analysis of the data, and writing and revising of the manuscript. **Mengying Li:** Formal analysis, Data curation, Writing – original draft, All authors contributed equally to this manuscript. All authors were involved in the planning of the research, analysis of the data, and writing and revising of the manuscript. **Hugo T.C. Pedro:** Formal analysis, Data curation, Writing – original draft, All authors contributed equally to this manuscript. All authors were involved in the planning of the research, analysis of the data, and writing and revising of the manuscript. **Carlos F.M. Coimbra:** Formal analysis, Data curation, Writing – original draft, All authors contributed equally to this manuscript. All authors were involved in the planning of the research, analysis of the data, and writing and revising of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank the U.S. Environmental Protection Agency and the South Coast Air Quality Management District (SCAQMD) for providing partial funding and on-site assistance. Additional thanks to Dr. Benjamin Kurtz from UC San Diego for his assistance with the cell modem setups, and to Dr. David P. Larson also from UC San Diego, for sensor development, installation and data management.

Appendix A. Evaluation metrics

Commonly applied statistical metrics [27,48] are used to evaluate the performance of irradiance derivation and forecast: the mean biased error (MBE, W/m^2),

$$\mathrm{MBE} = \frac{1}{K}\sum_{k=1}^{K}(\hat{I}_k - I_k), \tag{.1}$$

the mean absolute error (MAE, W/m^2),

$$\mathrm{MAE} = \frac{1}{\mathrm{k}}\sum_{k=1}^{k}\|\widehat{I}_k - I_k, \tag{.2}$$

the root mean square error (RMSE, W/m^2):

$$\mathrm{RMSE} = \sqrt{\frac{1}{K}\sum_{k=1}^{K}(\hat{I}_k - I_k)^2}, \tag{.3}$$

and the relative root mean square error (rRMSE,%),

$$\mathrm{rRMSE} = \frac{\mathrm{RMSE}}{1\Big/K\sum_{k=1}^{K}|I_k|}, \tag{.4}$$

In the above equations, K is the number of time instance in the dataset, $\hat{I}_k$ and I_k are the modeled and measured irradiance at time t_k, respectively.

References

[1] IEA, Technology Roadmap Solar Photovoltaic Energy, Tech. rep., IEA, 2014.
[2] California solar statistics, URL, www.californiasolarstatistics.ca.gov, 2015.
[3] R.H. Inman, H.T.C. Pedro, C.F.M. Coimbra, Solar forecasting methods for renewable energy integration, Prog. Energy Combust. Sci. 39 (6) (2013) 535–576.
[4] M.A. Ortega-Vazquez, D.S. Kirschen, Economic impact assessment of load forecast errors considering the cost of interruptions, in: Power Engineering Society General Meeting, 2006. IEEE, IEEE, 2006, p. 8.
[5] G.K. Singh, Solar power generation by PV (photovoltaic) technology: a review, Energy 53 (2013) 1–13.
[6] M.A. Martínez, J.M. Andújar, J.M. Enrique, A new and inexpensive pyranometer for the visible spectral range, Sensors 9 (6) (2009) 4615–4634.
[7] A. Habte, S. Wilcox, T. Stoffel, Evaluation of Radiometers Deployed at the National Renewable Energy Laboratory's Solar Radiation Research Laboratory, Tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States), 2015.
[8] M. Sengupta, A. Habte, S. Wilbert, C. Gueymard, J. Remund, Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications, Tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States), 2021.
[9] B. Kurtz, J. Kleissl, Measuring diffuse, direct, and global irradiance using a sky imager, Sol. Energy 141 (2017) 311–322.
[10] G. Wang, B. Kurtz, J. Kleissl, Cloud base height from sky imager and cloud speed sensor, Sol. Energy 131 (2016) 208–221.
[11] P. Pawar, C. Cortés, K. Murray, J. Kleissl, Detecting clear sky images, Sol. Energy 183 (2019) 50–56.
[12] A. Aides, A. Levis, V. Holodovsky, Y.Y. Schechner, D. Althausen, A. Vainiger, Distributed sky imaging radiometry and tomography, in: 2020 IEEE International Conference on Computational Photography (ICCP), IEEE, 2020, pp. 1–12.
[13] T. Fa, W. Xie, Y. Wang, Y. Xia, Development of an all-sky imaging system for cloud cover assessment, Appl. Opt. 58 (20) (2019) 5516–5524.

[14] R. Román, M. Antón, A. Cazorla, A.d. Miguel, F. Olmo, J. Bilbao, L. Alados-Arboledas, Calibration of an all-sky camera for obtaining sky radiance at three wavelengths, Atmos. Meas. Tech. 5 (8) (2012) 2013–2024.
[15] K. Tohsing, M. Schrempf, S. Riechelmann, H. Schilke, G. Seckmeyer, Measuring high-resolution sky luminance distributions with a ccd camera, Appl. Opt. 52 (8) (2013) 1564–1573.
[16] M. Kocifaj, H.A. Solano Lamphar, F. Kundracik, Retrieval of garstang's emission function from all-sky camera images, Mon. Not. Roy. Astron. Soc. 453 (1) (2015) 819–827.
[17] A. Hänel, T. Posch, S.J. Ribas, M. Aubé, D. Duriscoe, A. Jechow, Z. Kollath, D.E. Lolkema, C. Moore, N. Schmidt, et al., Measuring night sky brightness: methods and challenges, J. Quant. Spectrosc. Radiat. Transf. 205 (2018) 278–290.
[18] E. Scolari, F. Sossan, M. Haure-Touzé, M. Paolone, Local estimation of the global horizontal irradiance using an all-sky camera, Sol. Energy 173 (2018) 1225–1235.
[19] N.E. Herrera-Carrillo, M. Rivero, E. Gómez-Ramírez, R. Loera-Palomo, Solar irradiance estimation based on image analysis, in: 2018 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), IEEE, 2018, pp. 1–6.
[20] H. Jiang, Y. Gu, Y. Xie, R. Yang, Y. Zhang, Solar irradiance capturing in cloudy sky days–a convolutional neural network based image regression approach, IEEE Access 8 (2020) 22235–22248.
[21] L. Valentín, M.I. Peña-Cruz, D. Moctezuma, C.M. Peña-Martínez, C.A. Pineda-Arellano, A. Díaz-Ponce, Towards the development of a low-cost irradiance nowcasting sky imager, Appl. Sci. 9 (6) (2019) 1131.
[22] J.O. Kamadinata, T.L. Ken, T. Suwa, Sky image-based solar irradiance prediction methodologies using artificial neural networks, Renew. Energy 134 (2019) 837–845.
[23] S.R. West, D. Rowe, S. Sayeef, A. Berry, Short-term irradiance forecasting using skycams: motivation and development, Sol. Energy 110 (2014) 188–207.
[24] B. Urquhart, B. Kurtz, J. Kleissl, Sky camera geometric calibration using solar observations, Atmos. Meas. Tech. 9 (9) (2016) 4279–4294.
[25] Y. Chu, M. Li, C.F. Coimbra, Sun-tracking imaging system for intra-hour dni forecasts, Renew. Energy 96 (2016) 792–799.
[26] R. Marquez, C.F.M. Coimbra, Intra-hour DNI forecasting methodology based on cloud tracking image analysis, Sol. Energy 91 (2013) 327–336.
[27] M. Li, Y. Chu, H.T. Pedro, C.F. Coimbra, Quantitative evaluation of the impact of cloud transmittance and cloud velocity on the accuracy of short-term dni forecasts, Renew. Energy 86 (2016) 1362–1371.
[28] P. Blanc, B. Espinar, N. Geuder, C. Gueymard, R. Meyer, R. Pitz-Paal, B. Reinhardt, D. Renné, M. Sengupta, L. Wald, et al., Direct normal irradiance related definitions and applications: the circumsolar issue, Sol. Energy 110 (2014) 561–577.
[29] A. Kazantzidis, P. Tzoumanikas, A.F. Bais, S. Fotopoulos, G. Economou, Cloud detection and classification with the use of whole-sky ground-based images, Atmos. Res. 113 (2012) 80–88.
[30] Y. Chu, H.T.C. Pedro, L. Nonnenmacher, R.H. Inman, Z. Liao, C.F.M. Coimbra, A smart image-based cloud detection system for intra-hour solar irradiance forecasts, J. Atmos. Ocean. Technol. 31 (2014) 1995–2007.
[31] A. Sözen, E. Arcaklıoğlu, M. Özalp, N. Çağlar, Forecasting based on neural network approach of solar potential in Turkey, Renew. Energy 30 (7) (2005) 1075–1090.
[32] R. Marquez, C.F.M. Coimbra, Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the nws database, Sol. Energy 85 (5) (2011) 746–756.
[33] S.X. Chen, H.B. Gooi, M.Q. Wang, Solar radiation forecast based on fuzzy logic and neural networks, Renew. Energy 60 (2013) 195–201, 0.
[34] B. García-Domingo, M. Piliougine, D. Elizondo, J. Aguilera, CPV module electric characterisation by artificial neural networks, Renew. Energy 78 (2015) 173–181, 0.
[35] Y. Chu, M. Li, H.T.C. Pedro, C.F.M. Coimbra, Real-time prediction intervals for intra-hour dni forecasts, Renew. Energy 83 (2015) 234–244.
[36] Y. Chu, H.T.C. Pedro, C.F.M. Coimbra, Hybrid intra-hour DNI forecasts with sky image processing enhanced by stochastic learning, Sol. Energy 98 (2013) 592–603.
[37] P. Ineichen, A broadband simplified version of the solis clear sky model, Sol. Energy 82 (2008) 758–762.
[38] D. Yang, C. Gu, Z. Dong, P. Jirutitijaroen, N. Chen, W.M. Walsh, Solar irradiance forecasting using spatial-temporal covariance structures and time-forward kriging, Renew. Energy 60 (2013) 235–245.
[39] M. Chelbi, Y. Gagnon, J. Waewsak, Solar radiation mapping using sunshine duration-based models and interpolation techniques: application to Tunisia, Energy Convers. Manag. 101 (2015) 203–215.
[40] I. Loghmari, Y. Timoumi, A. Messadi, Performance comparison of two global solar radiation models for spatial interpolation purposes, Renew. Sustain. Energy Rev. 82 (2018) 837–844.
[41] M. David, Geostatistical Ore Reserve Estimation, Elsevier, New York, 1977.
[42] S. Rehman, S.G. Ghori, Spatial estimation of global solar radiation using geostatistics, Renew. Energy 21 (3) (2000) 583–605.
[43] G.K. Robinson, That BLUP is a good thing: the estimation of random effects, Stat. Sci. (1991) 15–32.
[44] A. Walther, A.K. Heidinger, Implementation of the daytime cloud optical and microphysical properties algorithm (DCOMP) in PATMOS-x, J. Appl. Meteorol. Climatol. 51 (7) (2012) 1371–1390, https://doi.org/10.1175/JAMC-D-11-0108.1.
[45] R. Perez, R. Seals, A. Zelenka, Comparing satellite remote sensing and ground network measurements for the production of site/time specific irradiance data, Sol. Energy 60 (2) (1997) 89–96, https://doi.org/10.1016/S0038-092X(96)00162-4.
[46] J.A. Ruiz-Arias, S. Quesada-Ruiz, E.F. Fernández, C.A. Gueymard, Optimal combination of gridded and ground-observed solar radiation data for regional solar resource assessment, Sol. Energy 112 (2015) 411–424, https://doi.org/10.1016/j.solener.2014.12.011. http://www.sciencedirect.com/science/article/pii/S0038092X14006021.
[47] M. Kocifaj, Unified model of radiance patterns under arbitrary sky conditions, Sol. Energy 115 (2015) 40–51.
[48] A. Kaur, L. Nonnenmacher, C.F. Coimbra, Net load forecasting for high renewable energy penetration grids, Energy 114 (2016) 1073–1084.