

Very short-term solar irradiance forecast using all-sky imaging and real-time irradiance measurements

M. Caldas^{a,*}, R. Alonso-Suárez^b

^a*Instituto de Física, Facultad de Ciencias, Universidad de la República. Iguá 4225, CP 11400, Montevideo, Uruguay.*

^b*Laboratorio de Energía Solar, Universidad de la República. Av. L. Batlle Berres, km 508, CP 50000, Salto, Uruguay.*

Abstract

A hybrid forecasting methodology to predict one-minute averaged solar irradiance one to ten minutes in advance is presented and evaluated. The methodology combines the use of all-sky images and irradiance measurements which are both processed in real time to produce the forecast. Pre-existing image processing techniques are locally adapted to estimate the mean motion of clouds, which is used to predict the future sun disk cover by clouds. Then, the predicted cloud information is converted into a solar irradiance estimate using the proposed model which uses real time measurements to extract its parameters for prediction. The validation of the method is done with a sample of 5238 forecasting time points, spread over a six-month period. The forecast uncertainty is assessed separately for clear, cloudy and partly cloudy days, revealing important characteristics of the model's performance under the different conditions. Under partly cloudy and highly variable conditions, positive forecasting skills with respect to regular persistence are achieved above forecasting horizons of two minutes, with a peak performance of 11.4% for forecasting horizons of six and ten minutes. The proposed model also outperforms a smart persistence model for all time horizons under these sky conditions. The model's ramp detection index (RDI, as defined in [Chu et al. \(2015\)](#)) is also evaluated for high and moderate ramps, achieving RDI indexes between 55 and 62% and between 43 and 49% for high and moderate ramps, respectively. These results show that in challenging highly variable solar irradiance conditions the proposed model is suitable for the very short term solar resource forecasting.

Keywords: All-sky images, solar irradiance forecast, ramp detection, cloud motion.

1. Introduction

The annual global installation of renewable power generation continues to increase year by year. Solar energy is not the exception and is taking an increasing role in the energy transition. Solar PV has shown record new installations in the last years, being the year 2016 the first one where the annual net additions were higher than any other renewable energy technology, and accounting by the end of 2017 of a new record high of more than 400 GW worldwide installed capacity ([REN21, 2018](#)). One of the major challenges associated with the increase of solar PV share into electricity grids is the variability of the solar input, which translates directly to fluctuations in the power output of solar PV plants ([Antonanzas et al., 2016](#); [Voyant et al., 2017](#)). While the variability due to Earth's rotation and translation is slow and predictable, the variability caused by the interaction between the solar radiation and the atmosphere (mainly clouds) is significantly higher and difficult to predict ([Kleissl, 2013](#)). These fluctuations affect the quality of the energy being generated and injected into the grid, mainly in the form of short term frequency and voltage instabilities ([Lave et al., 2013](#)).

Mitigation actions can be taken to deal with these effects, including geographical dispersion of PV plants ([Hoff and Perez, 2010](#); [Lave et al., 2012](#); [Marcos et al., 2011](#); [Rowlands et al., 2014](#)), usage of energy storage systems (ESS) ([de la Parra et al., 2015](#); [Senjyu et al., 2008](#)) and making use of forecasting techniques to anticipate the variations in the solar input. Unlike the former, the latter does not alter the power fluctuations themselves, but provides valuable information for grid operators in order for them to make suitable decisions to minimize its impact. The first step in predicting the future behavior of solar PV plants' output power is to predict global horizontal solar irradiance (GHI) at ground level.

There are different techniques for solar energy forecasting suitable for different temporal and/or spatial resolutions. They may be grouped according to the input data being used, which usually determines the forecast horizon range of the method

*Corresponding author: M. Caldas, mcaldas@fisica.edu.uy

27 (Diagne et al., 2013): statistical models based on real-time irradiance measurements, suited for the very short term (5 minutes
28 to some hours ahead); satellite imagery used to obtain large scale cloud motion vectors, suited for forecast horizons between 30
29 minutes up to 6 hours; numerical weather prediction (NWP) models suited for 1 hour onwards; and ground-based sky images
30 used also to obtain cloud motion estimates but with a higher spatial (sub-km) and temporal (30 to 60 seconds) resolution as
31 compared to satellite imagery. The latter is the focus of this work. The very short term forecasting is aimed to anticipate the
32 sudden variations of solar irradiance caused by clouds, namely the ramp events, reducing the needs for reserves or ancillary
33 services to compensate these quick variations and, thus, reducing imbalance problems and operational costs (Chow et al.,
34 2015; Ela et al., 2013).

35 Ground-based all-sky images have been used for solar irradiance forecasting in different ways. Marquez and Coimbra
36 (2013) and Chu et al. (2013) use a Total Sky Imager (TSI) for short-term forecasts (from 3 to 15 minutes ahead) of Direct
37 Normal Irradiance (DNI), by predicting the cloud fraction of the solar disk. In Marquez and Coimbra (2013) four days with
38 cumuliform clouds were selected to validate a technique based on a Particle Image Velocimetry (PIV) algorithm for the Cloud
39 Motion Field (CMF) estimation from the all-sky images. From the most significant direction, obtained from the CMF by
40 applying the k -means clustering algorithm, a ladder is constructed from the sun's position in where the cloud fractions X_i are
41 computed. Cloud fractions are used to forecast DNI using a very simplified model in where the factor $(1 - X_i)$ modulates the
42 maximum DNI value. In Chu et al. (2013) a generalized version of this DNI deterministic forecasting model is compared
43 against a Neural Network approach, obtaining better results with the latter. The evaluation was done over a testing set of 7500
44 forecasting points, extracted from 19 days in seasons with different low and high variability. In Quesada-Ruiz et al. (2014) an
45 adjustable-ladder method is introduced for calculating the CMF, providing better results for DNI forecast in comparison with
46 the PIV algorithm. The evaluation is done using 14 selected days from a 6 month period. More recently, Nou et al. (2018)
47 use high dynamic range (HDR) all sky images to predict DNI through a deterministic model and an Adaptive Network-based
48 Fuzzy Inference System (ANFIS). The validation is done over 80 days of data from July to October, 2015. In Chow et al.
49 (2011) a GHI (not DNI) all-sky forecasting method is proposed. The Cross-Correlation Method (CCM) is used to derive the
50 CMF from the all-sky images which is then used to forecast the next cloud cover image. A simplified model is used to forecast
51 irradiance from the cloud cover image as it is assumed that when the sun is covered by clouds, the GHI is 40% of the clear sky
52 value. This methodology is validated over 1920 forecasting points that correspond to the periods between 09:00 and 17:00
53 local times for 4 selected days from 3 months. Yang et al. (2014) use a sky imager to forecast GHI (from 5 to 15 minutes
54 ahead) on another location, relatively close to (less than 2 km) but not collocated with the imager, by generating a shadow
55 map at ground level and quantifying the impact of clouds covering the sun through statistical information of the clear-sky
56 index, obtained from a moving 2-hours window prior to the forecast. The performance assessment was done over 13000
57 forecasting points from 31 consecutive days during winter time where clear sky conditions, stratus and stratocumulus clouds
58 and partly cloudy conditions with low level cumulus, mid-level altocumulus and high-level cirrus clouds were included. In
59 Fu and Cheng (2013), the GHI is predicted by relating the clearness index (GHI normalized by the extraterrestrial horizontal
60 irradiance) with several features extracted from the time (t) all-sky images via linear regression. The performance evaluation
61 is done using a 4 week data set of a single month. None of the above mentioned works that use the so-called ladder method
62 have used information extracted from time series. Also, they either predict co-located DNI or GHI at another site different
63 from the location of the sky imager being used. Here, we assess the joint methodology to predict co-located GHI using a
64 proposed model for the combination of both sources of information.

65 In this work, we present a hybrid model for the combination of real-time irradiance measurements and all-sky images in
66 order to generate a 1 to 10 minutes-ahead forecast of one-minute averaged GHI in a collocated site. This model relates the
67 predicted clear-sky index with the predicted cloudiness by modulating its impact through two values obtained from real-time
68 measurements, which aim to quantify the effect of a covered and uncovered sun in the GHI time-series. The information
69 extracted from time-series is inspired in Yang et al. (2014), however, only two parameters from a 1-minute rate time-series
70 are extracted in our methodology (not three from a 30-seconds rate time-series) and following a different procedure. The
71 predicted cloudiness required for the model is obtained from the all-sky images using the so called ladder method with some
72 adaptations in the motion field filtering stage, used by Marquez and Coimbra (2013) and Chu et al. (2013) to forecast DNI.
73 Neither Marquez and Coimbra have used time-series information nor Yang et al. the ladder method, which is assessed by the
74 model presented here. The uncertainty of the system is evaluated over a 22 days data set where challenging highly variable
75 solar irradiance conditions were specifically selected. A detailed description of this data set is provided, in terms of a complete
76 list of the prevailing cloud types during the considered days as well as information on the data sets' statistics, which is rarely
77 found. The results are presented separately for different cloud conditions, for instance, partly cloudy, clear-sky and overcast,
78 with special focus in the former, where difficult cloud scenes were present in the all-sky images. By discriminating by day
79 type, different performance characteristics between them arise, as highlighted in this work. The testing site is located in the

80 south of South America (a region known as Pampa Húmeda), where solar irradiance forecast evaluation is scarce.

81 This article is organized as follows. Section 2 describes the GHI data acquisition and quality as well as the all-sky imaging
82 system being used, and a description of the data set. Section 3 describes the methodology to estimate cloud motion from
83 all-sky images, and how this is used to determine to what extent the solar disk will be occluded by clouds. This information is
84 combined with an analysis of the past 5 minutes of irradiance data, to provide a 1 to 10 minutes-ahead forecast of the clear-sky
85 index, which is then translated to GHI forecast by using the modified Kasten clear-sky model (Ineichen and Perez, 2002;
86 Kasten, 1980). The metrics being used to quantify the performance of the proposed methodology are described in Section 4,
87 along with the main results obtained on a sample of 5238 forecast time-points (3822 correspond to partly cloudy conditions).
88 Conclusions and proposed future improvements are contained in Section 5.

89 2. Data

90 An all-sky camera records images of the sky as seen from the ground with a typical sampling rate of one image per
91 minute, which is appropriate for cloud tracking and short-term solar irradiance forecasting. In addition to the all-sky images,
92 high quality solar irradiance measurements are also required in this work to implement in real-time the image-to-irradiance
93 model and to evaluate the forecast performance. Both data were recorded at an experimental research facility of the Solar
94 Energy Laboratory (LES, <http://les.edu.uy>), near the city of Salto, Uruguay. This facility is located in the northwest part
95 of the country, latitude -31.28°S and longitude -57.92°W , in a semi-rural environment with a sub-tropical temperate climate.

96 2.1. Image acquisition

97 Image acquisition is done by a locally designed all-sky system that is located at 10 meters from the solar irradiance
98 measurement equipment. The core of the system is an off-the-shelf DSLR camera (18 megapixel CMOS sensor) with a fish-
99 eye lens of equisolid angle projection (8mm f/3.5, full frame) attached to it, which provides a 180° field of view (FOV) in
100 the horizontal direction. The optical system is enclosed in a metallic box with a polycarbonate dome on top of the lens.
101 The enclosure contains temperature and humidity sensors as well as active elements to prevent the formation of dew and
102 frost on the dome. These sensors are controlled by an Arduino device, which is connected to a local computer. The images
103 are stored locally in this computer and are sent by FTP to a central web server that allocates the system’s webpage (<http://symphony.les.edu.uy/>),
104 from where the images are publicly available. Images are acquired at a resolution of 1920×1280
105 pixels and are registered with a frequency of one per minute. This locally build sky imager is shown in Figure 1-(a).

106 2.2. Solar irradiance measurements

107 Solar irradiance is recorded using a Kipp & Zonen Solys2 ground station (see Figure 1-(b)), which measures global
108 horizontal irradiance (GHI), along with the other solar radiation components. The GHI measurement is obtained using a Kipp
109 & Zonen CMP11 Secondary Standard pyranometer, that is calibrated every two years in our laboratory in accordance with
110 the ISO 9847:1992 standard. The calibration is done using a Kipp & Zonen CMP22 Secondary Standard that is kept with
111 traceability to the World Radiometric Reference (WRR, World Radiation Center, Davos, Switzerland). The measurement is
112 equipped with ventilation units which protect the pyranometer’s dome from dew and frost, specially during the morning, and
113 reduce the effect of raindrops and dust. The data are recorded at 1 minute interval as an average of four samples taken every
114 15 seconds by a factory calibrated Fisher-Scientific DT85 datalogger. As a part of the daily maintenance, the pyranometer’s
115 dome is regularly cleaned. The data recorded in this station is the highest quality solar radiation data set available in the
116 country and, considering the GHI equipment and maintenance, we evaluate its daily uncertainty at 3%. We have applied to
117 this data set the Baseline Solar Radiation Network (BSRN, <https://bsrn.awi.de>) quality criteria for physically impossible and
118 extremely rare GHI values (Yang et al., 2018), and less than 0.01% of the diurnal data was rejected.

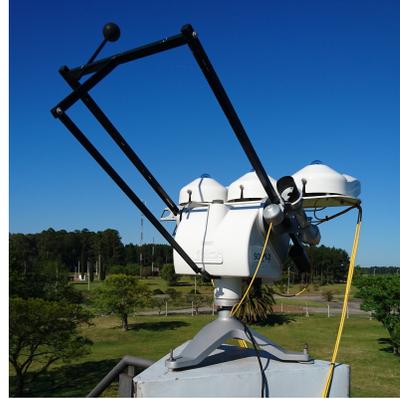
119 The GHI measurements, G_h , are normalized by an estimated clear-sky irradiance, G_h^{csk} , which results in the clear-sky
120 index, k_c , defined as,

$$k_c = \frac{G_h}{G_h^{\text{csk}}}. \quad (1)$$

121 In a similar manner as in Chow et al. (2011); Chu et al. (2013); Yang et al. (2014), the modified Kasten clear-sky model
122 (Ineichen and Perez, 2002; Kasten, 1980) is used to estimate the clear-sky irradiance in Eq. (1). The performance of this
123 clear-sky model has been evaluated over the region with an overall uncertainty of 6% (relative root mean square deviation,
124 rRMSD) using annual daily cycles of Linke turbidity factors, T_L , averaged over a set of years (Laguarda and Abal, 2017).
125 In this work, the required T_L values are obtained operationally from the previous days GHI time-series via a non-linear least



(a) Locally build all-sky imager.



(b) Kipp & Zonen Solys2 equipment.

Figure 1: Data acquisition equipment for solar irradiance measurement and all-sky images installed in the LES facility.

square regression (Levenberg, 1944; Marquardt, 1963) using the available clear-sky sub-series. The value obtained from the closer past day with at least 25% of clear-sky samples within ± 3.5 hours from solar noon is used. Thus, the system keeps track of the past T_L values and uses the last estimated value for the present day. With this procedure for calculating T_L , the clear sky model present a rRMSD of 2.5% with a small underestimation bias of -0.6% (relative mean bias deviation, rMBD), evaluated over the clear sky subset used in this work. It shall be noted that other daily T_L estimations may be used, such as locally derived average annual T_L cycles (Remund et al., 2003), simpler pre-adjusted empirical clear-sky parametrizations or other operational clear-sky irradiance models, such as the ESRA model (Rigollier et al., 2000) or the McClear model (Lefèvre et al., 2013), among others. A better turbidity estimation that the scheme used here may result in a better forecasting skill.

2.3. Data set description

To develop the proposed forecasting technique a set of coincident one-minute solar irradiance measurements and all-sky images is needed. For this study, information recorded between 12:00 and 16:00 (local time) for 22 different days from May 2016 to November 2016 was used. This sample is comparable to other samples in similar works, in terms of number of days, seasonal span and forecasting points, as detailed in Section 1. The selected days are specified in Table 1, grouping them into three categories: (i) partly cloudy days, (ii) clear-sky days and (iii) overcast days. The partly cloudy days were selected by seeking persisting variable sky conditions during the considered time span, avoiding significant subsets of clear sky or overcast conditions, and visually inspecting the cloudiness in the images, prioritizing different types of clouds and heights in the same scene. As shown in Table 1, days with only one or two cloud types represent a small fraction of the partly cloudy data set. Also, only in five days there is a portion of the time with clear-sky or overcast sky conditions. The whole data set is completed by selecting three random clear-sky and overcast days. The performance of the system does not vary much within different days under these sky conditions, thus, there is no point in adding more days for a separate assessment. The mean clear-sky index (μ_c) and the 1-minute variability (σ_c) of each day is included in Table 1, as well as the average for each sky condition. Here, the variability is calculated as the standard deviation of the 1-minute changes, $\sigma_c = \text{Std}\{\Delta k_c\}$ where $\Delta k_c = k_c(t) - k_c(t-1)$, following Hoff and Perez (2010) proposal. Both μ_c and σ_c were calculated using the considered sub-set of each day (from 12:00 to 16:00, local time). The limitation on local time span was due to the fact that the active elements of the all sky system used to prevent the formation of dew and frost on the dome was not operative on several days of the data set, leading to the presence of water droplets on the dome in the morning hours. This, in turn, affected the cloud detection accuracy, and so the starting time of 12.00 local time was chosen to guarantee an adequate image quality in the whole data set. Also, although it is possible to perform the forecast at each sample in less than 1 minute, the large number of samples to process pose a limit on the computational time required for the test. The total number of forecast time points is 5238 (698 clear sky, 718 overcast and 3822 partly cloudy). It must be noted that this distribution is not representative of long-term sky conditions for our location, since we want to make our performance evaluation as independent as possible from local conditioning. In terms of solar zenith angle the data spans from 8° to 70° .

An outline of the selected data set is presented in Figure 2, where σ_c is plotted against μ_c . The resulting chart is similar to the one proposed by Lauret et al. (2016), in which clear-sky, cloudy and partly cloudy days may be identified. The partly

Date	Sky condition	μ_c	σ_c	Cloud type(s)
17/05/2016	Partly Cloudy	0.82	0.17	Cu, Cu and Ac, Cu and CC, Cc, Sc, Sc and Cc, Cc and Ci, Cs
21/05/2016	Partly Cloudy	0.98	0.19	Cu, Cu and Cc, Cu and Sc
02/06/2016	Partly Cloudy	1.01	0.16	Cc, Ci, CLR, Cu
05/06/2016	Partly Cloudy	0.75	0.19	Cu and Sc, Sc, Sc and Cs
21/06/2016	Partly Cloudy	1.03	0.20	CLR-EF, Cu
06/07/2016	Partly Cloudy	0.91	0.20	Sc, Cu
15/07/2016	Partly Cloudy	0.54	0.19	Cc and Cu, CC and Sc, Sc, OVC
18/07/2016	Partly Cloudy	0.71	0.15	Cc, Cc and Cu, Cc and Sc
07/09/2016	Partly Cloudy	0.78	0.24	Cu, CLR-EF, Sc
13/09/2016	Partly Cloudy	0.86	0.42	Cu, Sc
29/09/2016	Partly Cloudy	0.90	0.14	Ci and Cc, Ac, CLR, Cc, Cc and Cs
30/09/2016	Partly Cloudy	0.66	0.26	Sc and Cs
06/10/2016	Partly Cloudy	0.75	0.24	Cu and Ci, Cu and Cs
07/10/2016	Partly Cloudy	0.44	0.15	Sc and Ci, OVC, Cc and Sc
26/11/2016	Partly Cloudy	0.70	0.18	Cs and Ac, Cs, Ac and Cu, Sc and Cs
29/11/2016	Partly Cloudy	0.94	0.29	Cu
Partly cloudy average		0.80	0.21	–
13/05/2016	Clear-sky	1.00	0.03	CLR
27/09/2016	Clear-sky	0.99	0.00	CLR
12/12/2016	Clear-sky	1.01	0.00	CLR
Clear sky average		1.00	0.01	–
27/05/2016	Overcast sky	0.23	0.02	OVC
26/07/2016	Overcast sky	0.24	0.04	OVC
24/09/2016	Overcast sky	0.21	0.01	OVC
Overcast sky average		0.23	0.02	–

Table 1: Description of the cloud types for each day in the data set. Indicated cloud types are: Cu, cumulus; Ac, altocumulus; Cc, cirrocumulus; Ci, cirrus; Sc, stratocumulus; Cs, cirrostratus. CLR, OVC and CLR-EF indicate clear-sky, overcast sky and clear-sky with prominent cloud evaporation and formation, respectively. Cloud types joined by "and" indicate simultaneous occurrence, as used in Yang et al. (2014). μ_c and σ_c are respectively the 1-minute average and variability in the considered 12:00 to 16:00 time span.

cloudy days are grouped in red: their daily one minute variability varies between 0.15 and 0.40 and their daily one minute mean value between 0.4 and 1.0. Clear-sky days, grouped in blue, present high k_c mean value and low variability, and are located in the right-bottom part of the diagram. Cloudy days present low k_c mean value and variability, and are grouped in orange in the left-bottom part of the diagram. It can be observed that clear sky days and overcast days are very close each other inside each cluster. The criteria used for selecting the partly cloudy days result in high variability days included in the sample. The different regions observed in Figure 2 support the separated evaluation performed in this work, as the conditions assessed are significantly different between each other. Figure 3 shows the probability distribution of the one-minute k_c records of the partly cloudy data set used in this work (in bars) in comparison with the distribution of the whole data set (in dashed line). The distribution of the whole data set peaks, as expected, at $k_c = 1$, since it includes clear-sky days. The values below the peak correspond to the partly cloudy and cloudy sky samples and the values above the peak correspond to measurements' overshoots usually found under partly cloudy conditions due to small clouds causing a high solar irradiance intermittence. The distribution for partly cloudy days show more percentage of values above $k_c = 1$, which are the overshoots and also less occurrences of low clear-sky index values ($k_c < 0.3$), as expected.

3. Methods

The three main steps of the forecasting technique are: (i) detection of clouds, (ii) cloud motion estimation and (iii) GHI prediction. The main contribution of this work is a deterministic model used to combine information extracted from all-sky images and real-time measurements, included in step (iii). The first two steps are based on pre-existing methodologies that were modified and/or adapted for our input information. The methodology that we use is explained in detail in the following.

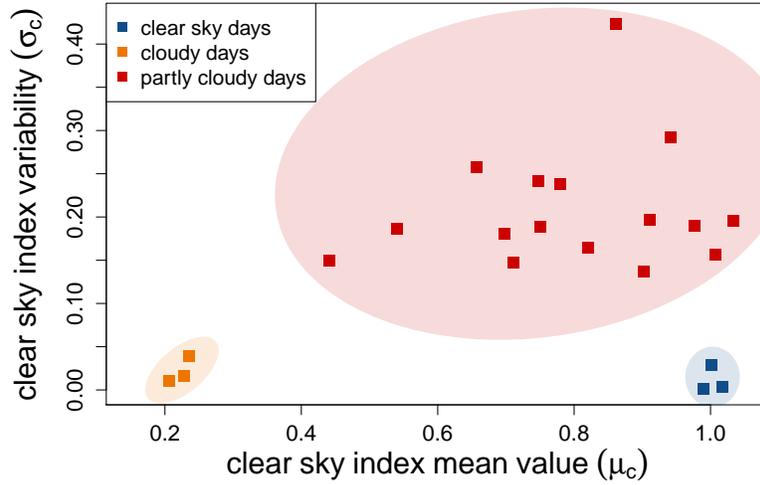


Figure 2: Clear-sky index mean value vs variability for the selected days.

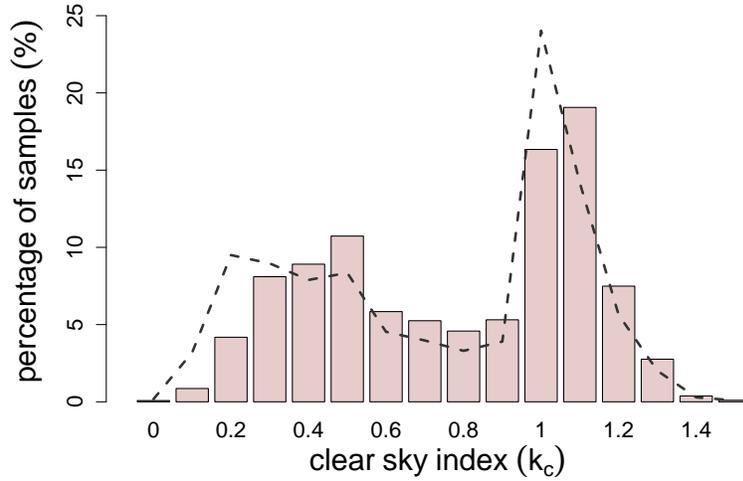


Figure 3: Clear-sky index values histogram of the selected data set. The dashed line corresponds to the distribution of the whole data set, including clear and overcast days. The bars correspond only to the partly cloudy data set.

178 The first step is an offline camera calibration procedure. This procedure is described in Section 3.1. Section 3.2 presents
 179 the cloud detection algorithm, which makes use of two well known techniques, namely the Fixed Threshold method (Li et al.,
 180 2011) and the Clear Sky Library (CSL) method (Ghonima et al., 2012) with an adaptive threshold, applied to the given image's
 181 red-to-blue ratio (RBR). Section 3.3 describes the method for estimating the mean cloud displacement. This is achieved by
 182 undoing the optical system's equisolid angle projection (through the calibration parameters described in Section 3.1) and
 183 applying the cross-correlation method (CCM, as described in Hamill and Nehrkorn (1993)) to a pair of consecutive projected
 184 images. After a series of post-processing steps applied to the obtained displacement field, a mean displacement vector is
 185 obtained. This mean vector is used to construct a ladder of grid elements originating in the sun's position and in the direction
 186 opposite to the mean displacement vector. This follows the idea described in Marquez and Coimbra (2013), where this ladder
 187 is used to forecast direct normal irradiance (DNI) via a deterministic model. Following this idea, we use the cloud fraction
 188 in each grid cell to forecast GHI through a deterministic real-time model described in Section 3.4. The result of the proposed
 189 method is a 1 to 10 minute GHI forecasting scheme. The proposed methodology requires, in addition to the all-sky images,
 190 real-time access to GHI time series of the past 5 minutes at least. A flow chart of the algorithm is shown in Figure 4.

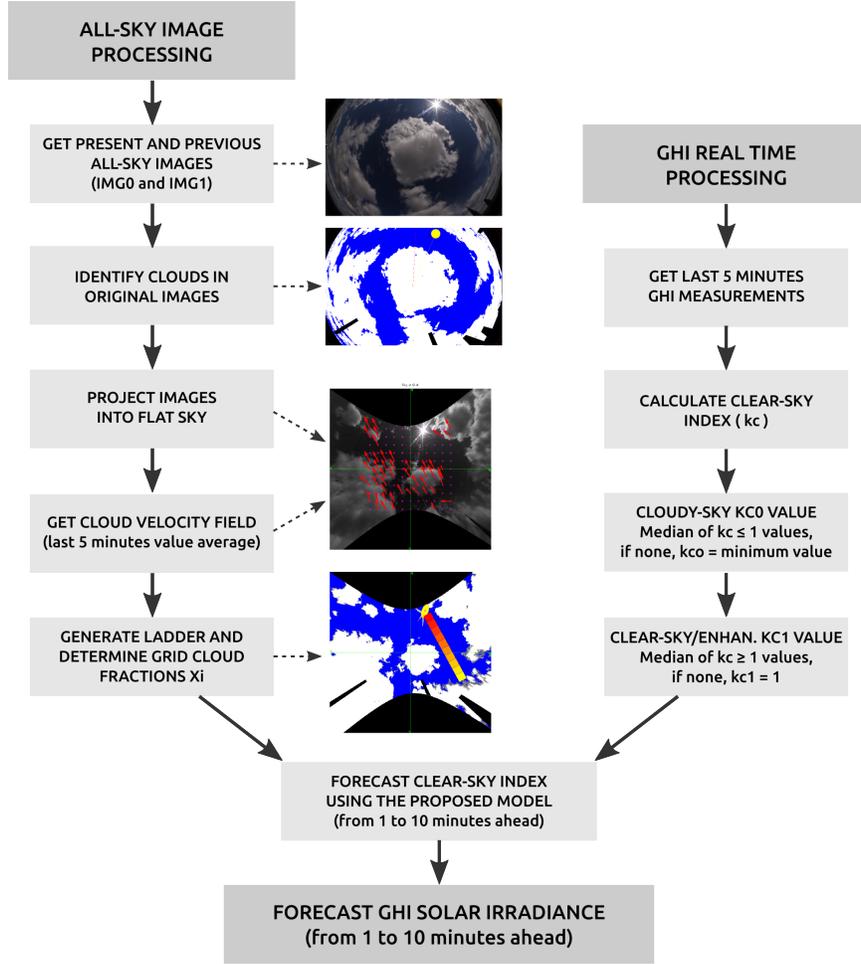


Figure 4: Flow chart of the main processing steps involved in the proposed method. As described in the text, the two input data are all-sky images and real-time GHI time series, which are used to forecast GHI one to ten minutes in advance.

191 3.1. Camera optical calibration procedure

192 The calibration of the camera's optical parameters, that is, to find the mapping functions between a pixel's position in the
 193 image to a position in the sky dome, can be done with the checkerboard method (Chauvin et al., 2015; Scaramuzza et al.,
 194 2006). Here we take another approach, similar to the one described in Urquhart et al. (2015), using the sun's position as a
 195 calibration tool. This calibration is done offline during the initial system configuration, stored locally and used afterwards
 196 as a part of the real-time forecasting system. First of all, the image's zenith (x_c, y_c) is estimated fitting a circle to a set of
 197 manually selected horizon points. The parameters of this fit are the circle radius (R) and center coordinates. Let (r, ϕ) be polar
 198 coordinates in the image Cartesian system (x, y) centered at (x_c, y_c) ; (θ, γ) the zenith and azimuth angles that locate a point in
 199 the sky dome, respectively; and p and s the sub-indexes that refer to pixel and sun, respectively. One of the main calibration
 200 parameters to be obtained is the angle for the North position with respect to the x axis, which we call α (see Figure 5).

Since we know that the mapping function of the lens being used is of equisolid angle type, we can find a linear fit of the form:

$$r = k_1 \sin\left(\frac{\theta}{2}\right) + k_2. \quad (2)$$

201 This assumes a perfectly vertically oriented camera, which is an approximation accurate enough for our purposes. The set
 202 of data points used in this fit were obtained by manually selecting the sun's center position (r_s, ϕ_s) on a set of non-occluded
 203 sun images. Calculating the sun's coordinates (θ_s, γ_s) , the linear fit shown in Figure 6-(a) was done in order to obtain the
 204 parameters k_1 and k_2 for Eq. (2). For our images these parameters were found to be $k_1 = 1417.4$ and $k_2 = -4.3$. The data
 205 set we use for this fit spans from 8° to 68° in solar zenith angles (θ_s) . Finally, γ_s and ϕ_s are used to find the orientation α

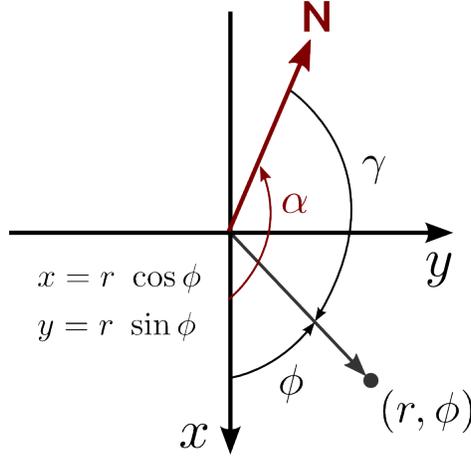


Figure 5: Angles involved in the azimuthal calibration of the camera.

206 of the North direction, through another linear fit based on Eq. (3), which express the relationship between these angles (the
 207 γ azimuth angle adapted for the southern hemisphere is defined here as $\gamma = 0$ for the North direction and positive/negative
 208 before/after solar noon). This fit is shown in Figure 6-(b), and resulted in an offset angle α of approximately 170° , which
 209 means that the North-South direction has only a 10° deviation.

$$\gamma = \alpha - \phi \quad (3)$$

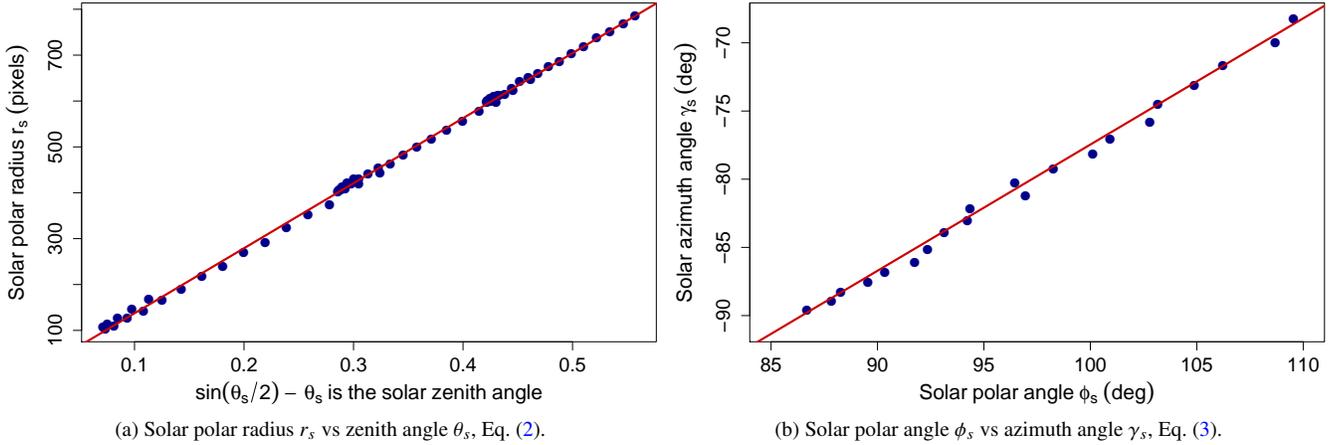


Figure 6: Linear adjustments for the offline optical calibration procedure.

210 Once the parameters of Eq. (2) and Eq. (3) are obtained offline using the calculated sun position (k_1 , k_2 and α), they can
 211 be used for every pixel in the image (r_p, ϕ_p) to calculate the positions (θ_p, γ_p).

212 3.2. Cloud detection algorithm

213 The first stage of the proposed cloud detection algorithm is to determine if the sun is occluded or not by optically thick
 214 clouds. In order to do this, the mean intensity within the sun disk area (from 0 to 255) is computed. The apparent sun size may
 215 vary depending on its position and atmospheric conditions. For the sake of simplicity, a fixed radius of 50 pixels is used. This
 216 fixed radius jointly with an adapted brightness threshold, allows us to gain simplicity in this step without losing significant
 217 accuracy in the cloud detection. Let's define the sun radius, R_s , as the average value, in the four x - y radial directions, of the
 218 smallest distance from the sun center in each direction at which the image intensity drops 5% from the value at the center.
 219 The mean brightness within this estimated radius and within a fixed 50 pixels radius was computed for the clear-sky images,

and is summarized in Table 2, which span from 10° to 70° in solar zenith angle. As shown, mean intensities of the solar disk, no matter if a fixed or dynamic radius is considered, are close to the saturation value. A maximum difference of 6.5% was obtained for the first day in the sample between the brightness in the two regions. The mean difference for the analyzed images was 2.2%. Thus, the mean brightness in a fixed radius sun disk shows small differences with more sophisticated schemes to define the radius R_s . Furthermore, when the sun is occluded by optically thick clouds the mean intensity drops significantly, typically below 200. Based on this analysis, a brightness threshold of 220 was set to distinguish between a clear and an occluded sun disk when using a fixed 50 pixels radius.

Date and time	Solar Zenith	Solar Radius	Mean intensity of solar disk		
	Angle (deg)	R_s (pixels)	Radius R_s	$R_s = 50$ pix.	Δ (%)
21/05/2016 15:54	68.2	34	248	232	-6.5
17/07/2016 15:59	56.5	24	254	240	-5.5
13/09/2016 14:03	39.1	44	248	251	+1.2
06/10/2016 12:20	26.2	34	252	249	-1.2
07/10/2017 15:09	43.2	36	252	250	-0.8
29/11/2018 12:27	10.1	37	250	249	-0.4

Table 2: Solar disk mean intensity for a radius R_s (computed as described in the text) and for a fixed 50 pixels radius.

The cloud detection algorithm is based on the image’s red-to-blue ratio (RBR). When the sun is occluded by optically thick clouds, a global fixed threshold differentiates accurately cloudy from clear pixels throughout the image, including the circumsolar region. This is because the image does not become close to saturation in the circumsolar region, which does occur when the sun disk is clear or occluded by optically thin clouds. The RBR threshold for distinguishing cloudy from clear pixels in the former situation was heuristically set to 0.9. No further post-processing of the decision image is needed when the sun is occluded by thick clouds. However, when the sky is occluded by very dark, optically thick clouds, in overcast sky situations, this threshold is too high and misclassification occurs, which occasionally leads to low cloud fractions. This has a negative effect on the system’s performance for forecasting solar irradiance in overcast sky situations, as will be discussed in Subsection 4.3. On the other hand, when the sun is not occluded by optically thick clouds, a global RBR threshold is not suitable because misclassification occurs, particularly in the circumsolar region. In this case, the RBR image is compared to a clear-sky image model, obtained from a clear-sky library (CSL) that was generated from our images in a similar fashion as described by Ghonima et al. (2012). Using only clear-sky images, a four degree polynomial clear-sky RBR model was adjusted to the RBR level across the image as a function of the pixel’s zenith angle, θ_p , and the sun-pixel-angle, ψ_{ps} , as,

$$\text{RBR}(\theta_p, \psi_{ps}) = \sum_{i=0}^{i=4} \sum_{j=0}^{j=4} p_{ij} \theta_p^i \psi_{ps}^j. \quad (4)$$

The angles θ_p and ψ_{ps} for this parametrization are specified in Figure 7. The coefficients p_{ij} were adjusted using several images, having approximately the same solar zenith angle (θ_s). A pixel in the given image is classified as cloudy if the difference between its RBR and the corresponding clear-sky RBR model is above a certain threshold. A fixed threshold $\delta_{\text{diff},0}$, obtained by applying Otsu’s method (Otsu, 1979) on the difference image, classifies pixels accurately in most parts of the image. However, since our main purpose is to develop a very short term solar forecasting model, a subtle correction is introduced to avoid misclassification in the closest vicinity of the sun, which affects the shortest forecast horizons of such a model. This correction is implemented through a ψ_{ps} dependence of the threshold, given by:

$$\delta_{\text{diff}}(\psi_{ps}) = \delta_{\text{diff},0} \left(1 + \exp \left\{ -\frac{\psi_{ps}}{5} \right\} \right), \quad (5)$$

with ψ_{ps} in degrees. This raised threshold in the closest vicinity of the sun is as pointed out a subtle correction that avoid misclassification of clear pixels as cloudy, mainly due to glare and reflection artifacts in this region. The value of the scaling factor in the exponential term (5 degrees) was tuned manually by visual inspection of the results obtained on cloud identification.

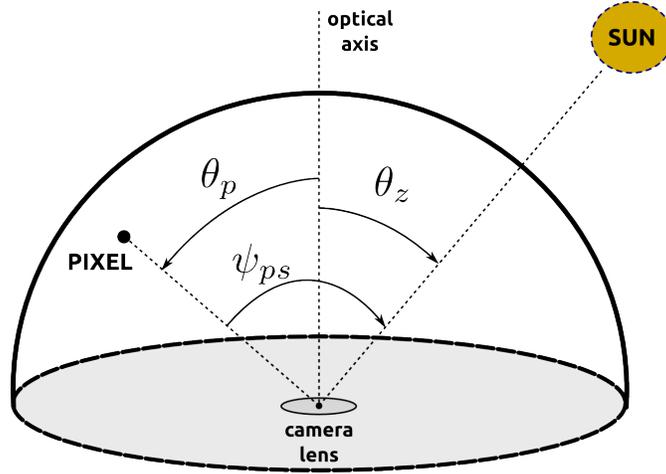


Figure 7: Angles θ_p and ψ_{ps} that are required for the clear-sky image model.

250 Once δ_{diff} is computed, a decision image is generated by classifying cloudy pixels as those having a difference above δ_{diff} .
 251 A final post-processing step on the decision image is applied, based on the total cloud fraction, CF, obtained from the classified
 252 image: if CF is below 0.05, all pixels in the region of interest are classified as clear, and the retrieved cloud fraction is 0. This
 253 is because spurious cloudy pixels are unavoidable in clear or near clear-sky images (due to glare and other reflection artifacts).
 254 This threshold on CF was selected based on the analysis of the subset of the images considered for this study.

255 Figure 8 shows decision images obtained for different sky conditions, using the CSL and fixed threshold methods. This
 256 sequence clearly shows that in the case when the sun disk is clear, the use of the CSL algorithm holds good results in the
 257 whole image, whereas the fixed threshold method misclassifies clear pixels as cloudy (mainly optically thin clouds, shown
 258 in gray) in the circumsolar region. On the other hand, when the sun disk is occluded by clouds, as in (d) and (j), the CSL
 259 method is not suitable due to the fact that the overall brightness and also RBR of the image is lowered as compared to a clear
 260 sun disk situation. This makes it unsuitable to compare such images with a CSL, which is based on clear sky conditions.
 261 For instance, in (e) a considerable clear sky portion of the image is classified as cloudy, while in (k) a circumsolar ring is
 262 misclassified as clear in an overcast sky condition. Thus, the proposed cloud detection algorithm makes use of the advantages
 263 of each method on either situation: if the sun disk is occluded, a fixed threshold is used to classify pixels, with the additional
 264 benefit of lower computational cost; if the sun disk is clear, the image is compared to a clear-sky image model obtained
 265 from our CSL. It should be noted that even though the proposed method improves significantly the results in the circumsolar
 266 region (which always poses a challenge for any cloud detection algorithm), in as much as solar forecasting is concerned the
 267 cloud detection in this region affects primarily the nearest future forecasting (say, 1 to 3 minutes ahead), for which persistence
 268 models are rarely outperformed. The final decision image is obtained as described previously by classifying pixels as cloudy
 269 if the difference is above the threshold δ_{diff} . Figure 9-(a) shows a clear-sky image model generated for identifying clouds on
 270 the image of Figure 8-(a), where the sun is not occluded by clouds, but clouds are present in its vicinity. Figure 9-(b) shows
 271 the difference between the given image's RBR and the model image, which is based on the polynomial fit of Eq. (4) for the
 272 given solar zenith angle. The correction of Eq (5) deals with the reflection artifacts observed in the sun's vicinity, yielding the
 273 decision image of Figure 8-(b).

274 3.3. Cloud displacement

275 The cloud motion field is computed from a pair of consecutive images using the CCM method proposed by Hamill and
 276 Nehrkorn (1993). This method consists of subdividing the past image into a grid of squared cells of fixed size. For each cell
 277 in the image, the normalized cross-correlation between this cell and the neighboring cells in a larger search area of the present
 278 image is computed. The direction corresponding to the largest cross-correlation coefficient is assigned to the cell, as it is an
 279 indicator of the cloud displacement in this subregion of the image. The method is applied on red channel images as described
 280 in Chow et al. (2011), by generating a centered square of size equal half the size of the image. This square is divided into a
 281 10×10 grid. The search distance of the CCM algorithm was fixed at twice the grid element size. Only the central part of the
 282 image is considered due to the fact that the CCM algorithm is computationally expensive and also to avoid errors introduced
 283 by the projection at large zenith angles. Following Chow et al. (2011), vectors with cross-correlation coefficients less than

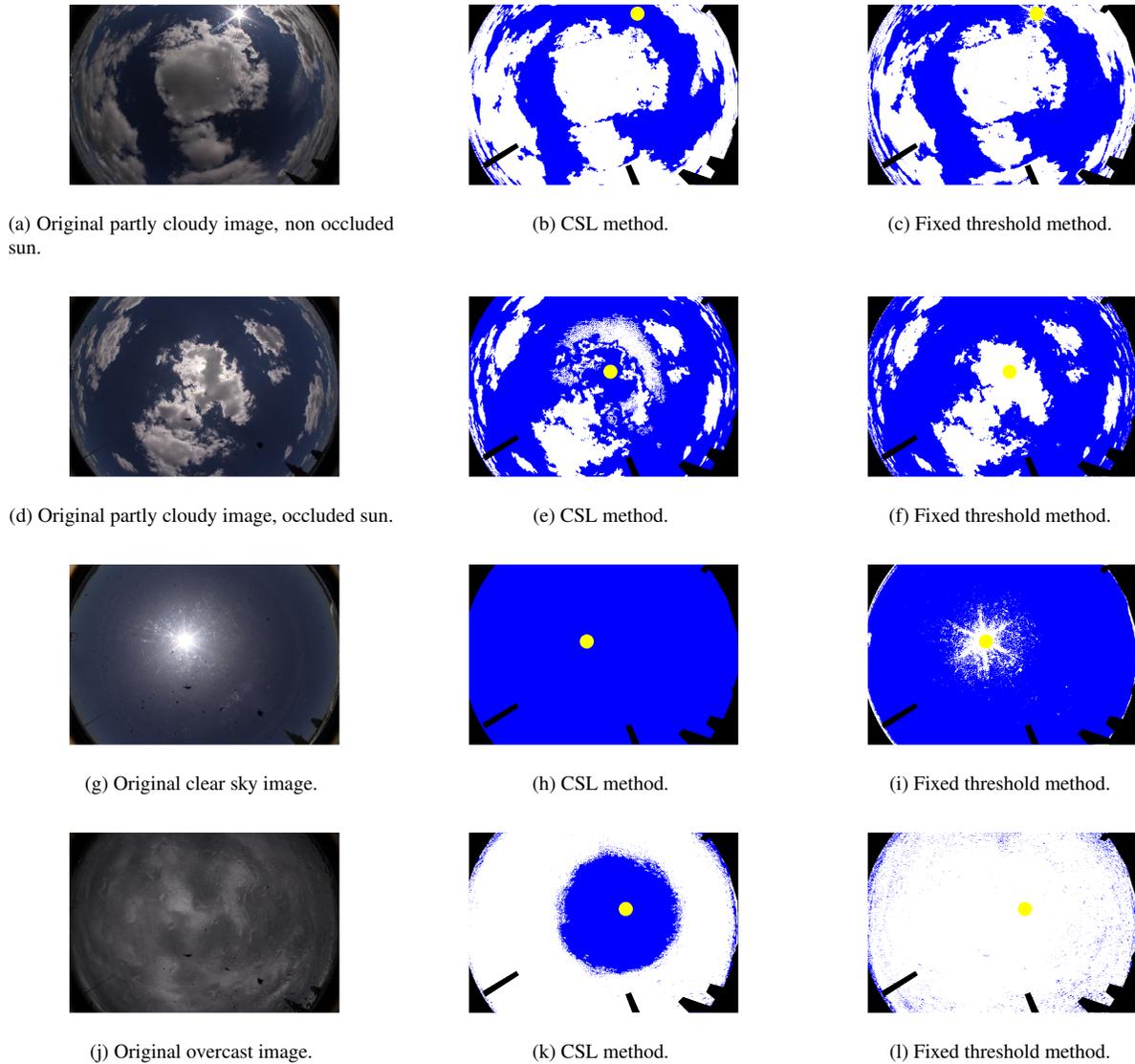


Figure 8: Results of cloud detection using either the CSL method with adaptive threshold (second column) or the fixed threshold method (third column), for different sky conditions: (a), (d) partly cloudy, (g) clear-sky, (j) overcast.

284 0.8 were discarded (see Figure 10-(a) and (b)). In [Chow et al.](#), several post-processing quality control procedures are used to
 285 ensure a suitable representation of the displacement, based on the mean and standard deviation of the velocity field. In our
 286 scheme, the resulting vector field is post-processed by considering histograms of speed and direction of the vectors. When
 287 only one cloud layer is present, the peak in each histogram is easily identified. When two or more cloud layers are present,
 288 peaks corresponding to each layer can be observed in the histograms. This has the advantage of being able to deal with
 289 multiple layers separately, instead of mixing the corresponding averages in one single value. The most frequent multilayer
 290 situation is that of two layers. In this case, the peak corresponding to the fastest moving layer is identified (lower altitude
 291 layer), as it has greater effect on very short-term irradiance variations. Once we have calculated this most probable speed
 292 and direction of the velocity field at the current time (see Figure 10-(c)), we apply a real-time 5-minute moving average filter
 293 in speed and direction to reduce the outliers effect in the motion time-series. The result of the moving average is the cloud
 294 displacement vector for the current time. Finally, following the methodology proposed by [Marquez and Coimbra \(2013\)](#) a grid
 295 with the form of a ladder is constructed in the opposite direction to the mean displacement vector (see Figure 10-(d)). Each
 296 grid cell has a constant width of 50 pixels, and a length equal to the mean cloud displacement in one minute obtained from

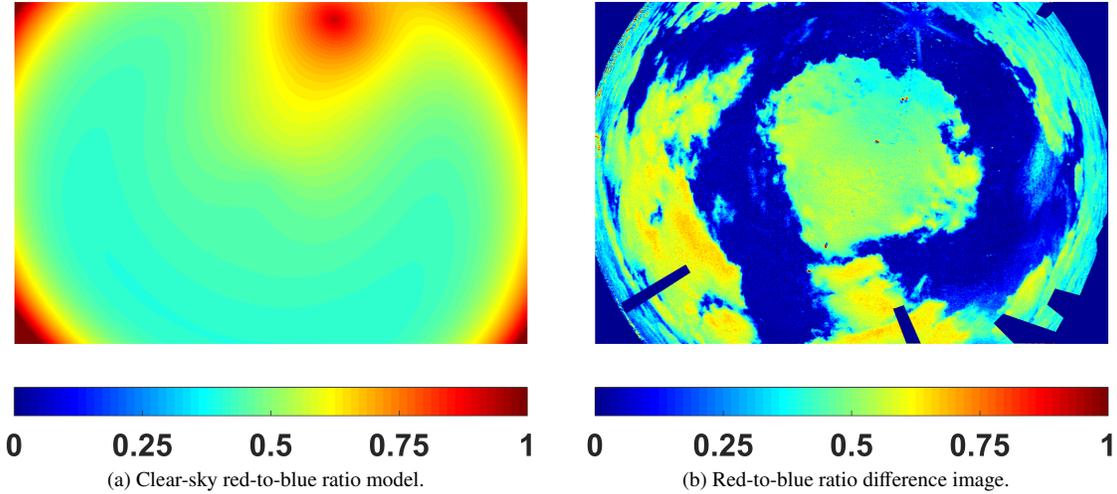


Figure 9: Images involved in the CSL method for identifying cloudy pixels on the cloudy image with non occluded sun of Figure-8-(a).

297 the mean speed. The total number of cells is fixed and set to 10, although a dynamic length could be a future improvement
 298 so that forecasts horizons longer than 10 minutes could be achieved, depending on mean cloud speed and orientation. The
 299 computation time for this algorithm on a regular computer (i7 processor, 3.4 GHz, 8 GB RAM) is of few seconds, which is
 300 compatible with a real-time forecasting system. After building this ladder, the cloud fraction X_i of the i -th cell is computed (i
 301 from 1 to 10), which are the image's inputs to the forecasting model as explained in the following Section.

302 3.4. GHI forecasting model and real-time processing

303 GHI at ground level is composed of direct irradiance arriving from the sun's disk, affected mainly by clouds covering the
 304 sun, and diffuse irradiance arriving from all directions, affected mainly by the clouds' distribution in the sky dome. For the
 305 sake of simplicity, in this first model proposal, the assumption that the GHI value is mainly affected (in the very short term) by
 306 clouds covering the sun is made. In this case, the overall effect on GHI due to variations in the relative position of the clouds
 307 is dominated by clouds covering the sun, particularly in as much as irradiance ramps are concerned.

308 The proposed model does not forecast GHI directly, but rather the clear-sky index k_c , defined in Eq. (1) as explained in
 309 Section 2.3. In a similar way as proposed in Yang et al. (2014), we inspect the past k_c time series to obtain information about
 310 the immediate past effect of the prevailing cloud type in the sky. In this proposal, two k_c values are obtained from the past 5
 311 minutes time-series: a k_c^{csk} value for a clear sun and a k_c^{occ} for an occluded sun. These values aim to represent the effect on the
 312 GHI solar irradiance of the prevailing cloud type and its distribution on the sky. With the grid cloud fractions X_i computed, as
 313 well as k_c^{csk} and k_c^{occ} , the proposed deterministic forecast model for the clear-sky index is given by,

$$k_c(t_0 + \Delta t_i) = k_c^{\text{csk}} + X_i \times (k_c^{\text{occ}} - k_c^{\text{csk}}), \quad (6)$$

314 where t_0 is the present time and Δt_i is the i -th forecast horizon, which varies between 1 and 10 minutes. If X_i equals 0, then
 315 the predicted k_c equals k_c^{csk} , and if X_i equals 1, then the predicted k_c equals k_c^{occ} . The values of X_i between 0 and 1 modulate
 316 the difference between the clear-sky and occluded sun values of k_c . Once k_c is predicted, it is straightforward to forecast G_h ,
 317 using the clear-sky model as,

$$\hat{G}_h(t_0 + \Delta t_i) = k_c(t_0 + \Delta t_i) \times G_h^{\text{csk}}(t_0 + \Delta t_i). \quad (7)$$

318 The clear-sky index that represents a clear sun, k_c^{csk} , is computed from the past time-series as the median of the $k_c > 0.9$
 319 values. As there may be overshoots in the past measurements, this is a way to partly consider them into the forecast. In
 320 a similar manner, a value that represents the effect of clouds covering the sun, k_c^{occ} , is computed as the median of the past
 321 $k_c \leq 0.9$ values. No significant difference was obtained when using the mean value instead of the median, so as the probability
 322 distributions of the samples are not symmetric, we chose the latter. If there are no k_c values above 0.9 in the time series being
 323 considered, k_c^{csk} is set to 1.0; if no values below 0.9 are found, k_c^{occ} is set to the minimum k_c in the time-series. The size
 324 of the k_c past moving window from which to extract the k_c^{csk} and k_c^{occ} values arise as a question. To quantify the effect of

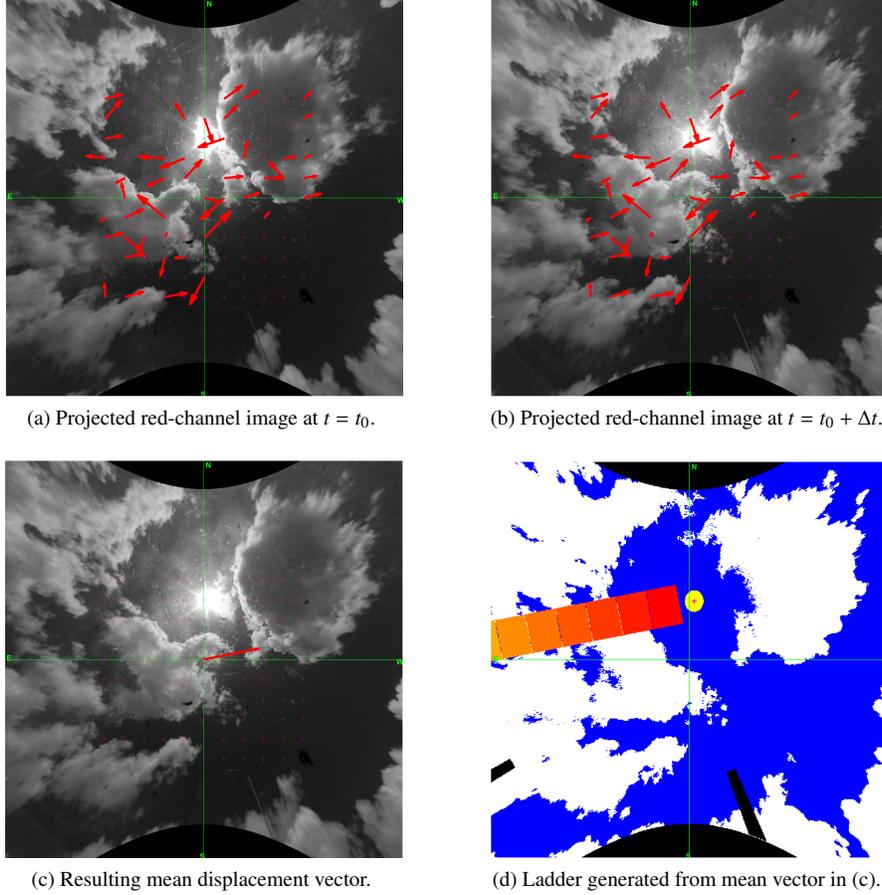


Figure 10: Results of the CCM algorithm for computing the mean cloud displacement. Vectors shown in (a) and (b) are those having $cc > 0.8$.

325 different window sizes on the methodology performance, we tested it on four partly cloudy days from Table 1 with different
 326 window sizes: 30, 15 and 5 minutes. We also evaluated a simplified method that uses fixed k_c^{csk} and k_c^{occ} values in 1.0 and 0.4
 327 respectively. The best results were achieved with the 5 minutes window (in terms of root mean squared deviation, RMSD,
 328 defined in Subsection 4.1). Thus, in this proposal, only the five past k_c values are considered for the real-time processing. Also,
 329 the worst results were achieved using the fixed k_c^{csk} and k_c^{occ} values, which demonstrates the benefit of extracting information
 330 from the past GHI time-series.

331 4. Results and discussion

332 This section presents the results obtained when using Eqs. (6) and (7) to forecast 1 to 10 minutes in advance the GHI at
 333 ground level. These results are presented in terms of well-known forecasting performance metrics, which are described in
 334 Subsection 4.1. The model is evaluated in absolute and relative terms, and its performance is compared against two clear-sky
 335 index persistence models. The first one uses the present k_c value as the persistence value, and so it is referred to as "regular
 336 persistence". The second one uses the average k_c in the same moving window used for the real-time processing (5 minutes), \bar{k}_c ,
 337 and it is referred to as "smart persistence". In addition, the proposed model is evaluated in its ability to forecast ramps in GHI.
 338 Subsections 4.2 and 4.3 present the main results for the partly cloudy data set, and clear and overcast data sets, respectively.

339 4.1. Performance metrics

340 Four metrics that are commonly employed to evaluate solar forecasting approaches and a ramp detection index proposed
 341 by Chu et al. (2015) are used. Three of them, the mean bias deviation (MBD), the mean absolute deviation (MAD) and the

342 root mean square deviation (RMSD), are calculated directly from the measured and predicted irradiances ($G_h(t + \Delta t)$ and
 343 $\hat{G}_h(t + \Delta t)$, respectively) as follows,

$$\text{MBD} = \frac{1}{n} \sum_{t=1}^n (\hat{G}_h(t + \Delta t) - G_h(t + \Delta t)), \quad (8)$$

$$\text{MAD} = \frac{1}{n} \sum_{t=1}^n |\hat{G}_h(t + \Delta t) - G_h(t + \Delta t)|, \quad (9)$$

$$\text{RMSD} = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{G}_h(t + \Delta t) - G_h(t + \Delta t))^2}. \quad (10)$$

344 The relative values for these metrics, rMBD and rRMSD, are expressed as a percentage of the mean measured irradiance,
 345 $\bar{G}_h(t)$. This set of metrics is computed both for the proposed model and the persistence models (regular and smart persistence),
 346 defined as,

$$\hat{G}_{hp}(t_0 + \Delta t_i) = k_{c,0} \times G_h^{\text{csk}}(t_0 + \Delta t_i), \quad (11)$$

347 where $k_{c,0} = k_c(t_0)$ for regular persistence and $k_{c,0} = \frac{1}{n+1} \sum_{t_0-n}^{t_0} k_c(t)$ for smart persistence (with $n = 5$ in this case). The MBD
 348 and RMSD metrics allow to compare the model's performance against both persistence types under the same solar irradiance
 349 data set, in particular, under the same irradiance variability.

350 A fourth metric, the forecasting skill (Coimbra et al., 2013), FS, can be calculated from the RMSD values as,

$$\text{FS} = 1 - \frac{\text{RMSD}_m}{\text{RMSD}_p}, \quad (12)$$

351 where RMSD_p and RMSD_m refer to the persistence and the proposed model, respectively. This metric expresses how well the
 352 model outperforms persistence (regular or smart) in terms of the RMSD metric.

353 Finally, we used a metric that is specifically suited for the very short-term forecast and is related to the model's ability to
 354 forecast ramps, namely the ramp detection index (RDI) as proposed by Chu et al. (2015). For a given forecast horizon Δt_i , the
 355 non-dimensional ramp magnitude (RM^*) is defined as the normalized difference between the present value of the irradiance
 356 and the irradiance at $t_0 + \Delta t_i$, relative to present time clear-sky irradiance:

$$\text{RM}^* = \frac{|G_h(t_0) - G_h(t_0 + \Delta t_i)|}{G_h^{\text{csk}}(t_0)}. \quad (13)$$

357 In this paper, as proposed by Chu et al., a high-magnitude ramps is considered when $\text{RM}^* \geq 0.5$ and a moderate ramp is
 358 considered when $0.3 < \text{RM}^* < 0.5$. A "hit" is defined when the absolute difference between actual value $G_h(t_0)$ and the
 359 predicted value $\hat{G}_h(t_0 + \Delta t_i)$ is greater than $0.1 \times G_h^{\text{csk}}(t_0)$, and the sign of the predicted difference is the same as the measured
 360 one. By counting the number of "hits", N_{hit} , and the total number of ramp occurrences, $N_{\text{hit}} + N_{\text{miss}}$, the ramp detection index
 361 is calculated as:

$$\text{RDI} = \frac{N_{\text{hit}}}{N_{\text{hit}} + N_{\text{miss}}}. \quad (14)$$

362 4.2. Results for partly cloudy days

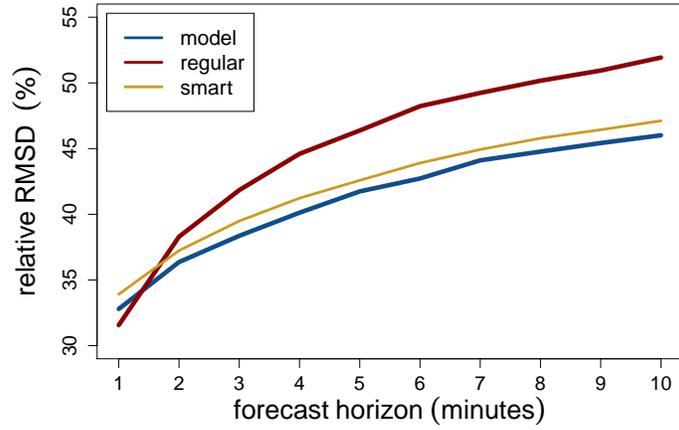
363 The results for the whole partly cloudy data set are shown in Table 3 and Table 4, and are summarized in Figures 11-(a) to
 364 (c). The absolute RMSD values for regular and smart persistence and the model reflect the challenging variability of the data
 365 set being used. Its high variability results in larger uncertainties as compared to other assessments made on mixed data sets,
 366 composed of all three classes of sky conditions. Further, the shorter time horizons do not necessarily present the typical low
 367 uncertainties observed for the latter, as it can be seen by inspecting the persistence performance on our data set. We obtained
 368 persistence rRMSD ranging from 31.5% to 51.9% for 1 to 10 minutes ahead respectively. For reference, another assessment
 369 of persistence made for this region in Giacosa and Alonso-Suárez (2018) achieved rRMSD of 17.6% to 32.9% for the same
 370 time horizons, but using a 5-years climate-representative data set that includes clear and overcast days (with low variability),
 371 thus lowering the rRMSD as compared to the data set used in this Section, composed only of partly cloudy days.

	MAD			MBD			RMSD		
	model	regular	smart	model	regular	smart	model	regular	smart
Δt_i	Wm ⁻²								
1 min	112	90	122	25.5	0.5	1.2	184	177	190
2 min	127	120	137	17.2	1.2	1.9	203	214	208
3 min	136	137	147	11.0	2.0	2.7	214	233	220
4 min	144	152	155	9.3	2.7	3.4	223	248	229
5 min	151	161	161	10.2	3.4	4.1	231	257	236
6 min	155	170	167	12.0	4.1	4.8	236	266	242
7 min	161	175	171	13.7	4.7	5.4	243	271	247
8 min	163	179	175	15.0	5.2	5.9	246	275	251
9 min	167	183	178	15.8	5.8	6.5	248	279	254
10 min	168	188	180	16.9	6.3	7.0	251	283	257
Δt_i	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
1 min	20.0	16.0	21.8	4.54	0.08	0.21	32.8	31.5	33.9
2 min	22.8	21.4	24.5	3.08	0.21	0.34	36.4	38.3	37.2
3 min	24.5	24.7	26.4	1.98	0.36	0.48	38.4	41.8	39.5
4 min	25.9	27.3	27.9	1.67	0.49	0.62	40.1	44.6	41.2
5 min	27.2	29.1	29.0	1.84	0.62	0.74	41.7	46.4	42.6
6 min	28.2	30.8	30.2	2.17	0.75	0.87	42.7	48.2	43.9
7 min	29.3	31.8	31.1	2.50	0.85	0.98	44.1	49.2	44.9
8 min	29.8	32.7	31.9	2.74	0.95	1.08	44.8	50.2	45.8
9 min	30.4	33.5	32.5	2.89	1.07	1.19	45.4	50.9	46.4
10 min	30.9	34.4	33.0	3.09	1.16	1.29	46.0	51.9	47.1

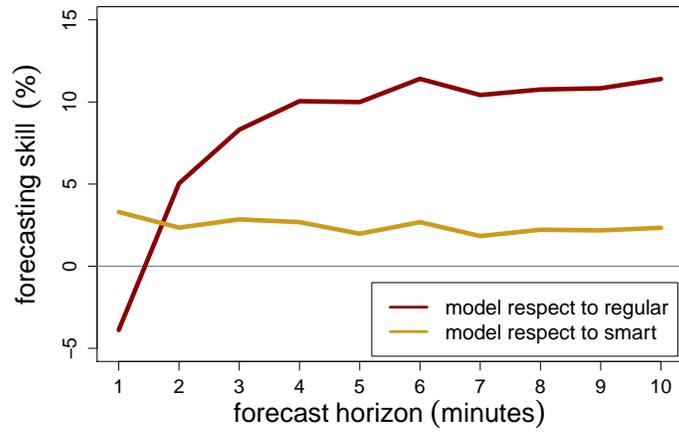
Table 3: Performance metrics of persistence and proposed model for partly cloudy days in absolute and relative terms.

372 The results show that the model outperforms smart persistence for all forecast horizons, although very moderately (FS
373 between 2–3%, approximately). Compared to regular persistence, the model starts outperforming it at $\Delta t_i = 2$ minutes. The
374 best performance in terms of forecasting skill is achieved at a forecast horizon of 6 and 10 minutes (FS = 11.4%). It has to be
375 noted that in contrast to the better performance in terms of RMSD, the model present higher mean bias in comparison with the
376 persistence procedures (that usually present very low or even negligible bias), tending to sub-estimate the solar irradiance for
377 all forecast horizons. In terms of the ability of the proposed model to correctly detect ramps, it is found that high-magnitude
378 ramps (HMR) are detected with a 61–62% accuracy for the first forecasting horizons (1 to 3 minutes ahead) and then the
379 performance reduces to about 55% accuracy for the last forecasting horizons (8 to 10 minutes ahead). For moderate ramps,
380 no clear performance tendency is found with the forecasting horizon, but rather it remains approximately constant at an RDI
381 between 43 and 49% accuracy. As expected, RDI increased with RM; high-magnitude ramps ($RM^* \geq 0.5$) are better predicted
382 than moderate ramps ($0.3 < RM^* < 0.5$), in agreement with the results obtained by [Chu et al. \(2015\)](#).

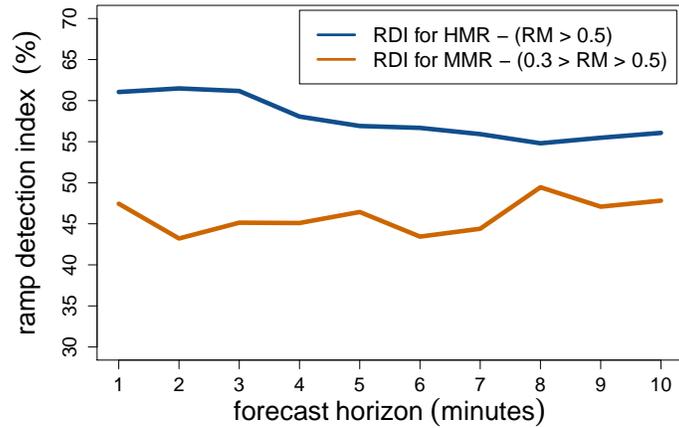
383 Different cloud types and their distribution in the sky significantly affect performance metrics, so it's not straightforward to
384 compare the results with other works without a detailed specification of the data set being used. However, some comparisons
385 can be done, mostly in terms of the FS, which, for instance, isolate the performance from the local solar radiation variability
386 ([Coimbra et al., 2013](#)). In [Chu et al. \(2015\)](#) a similar FS as the one reported here is obtained for the GHI 10-minutes forecast



(a) Relative root mean squared deviation (rRMSD).



(b) Forecasting Skill (FS).



(c) Ramp Detection Index (RDI).

Figure 11: Performance metrics for the whole partly cloudy data set and for the 21st May 2016: (a): rRMSD of persistence (orange) and proposed (blue) model; (b): forecasting skill; (c) RDI for HMR (blue) and MMR (orange).

387 (FS $\approx 10\%$), for one site with a 10-minutes variability of $\sigma_{10\text{-min}} = 0.11^1$, which is significantly smaller than in our partly

¹Defined as the standard deviation of the clear-sky index changes in a 10-minutes basis. Note that this is not the variability presented in Table 1 and

	RDI		Δt_i	RDI	
	HMR	MMR		HMR	MMR
Δt_i	(%)	(%)		(%)	(%)
1 min	61.0	47.4	6 min	56.7	43.4
2 min	61.5	43.2	7 min	55.9	44.4
3 min	61.2	45.1	8 min	54.8	49.4
4 min	58.1	45.1	9 min	55.5	47.1
5 min	56.9	46.4	10 min	56.1	47.8

Table 4: Detection metrics for High Magnitude Ramps (HMR) and Moderate Magnitude Ramps (MMR), for different forecasting horizons Δt_i .

cloudy data set ($\sigma_{10\text{-min}} = 0.24$). For a site with $\sigma_{10\text{-min}} = 0.15$ a FS of about 6% is reported. In [Yang et al. \(2014\)](#) a detailed description of the data set being used is included, with cloud types and FS for each day along with performance metrics for the whole data set. However, since the methodology is used to forecast GHI on a non collocated site, the results are not directly comparable to the ones reported here. [Yang et al.](#) report small negative forecasting skills for 5 minutes ahead forecast, when dealing with the added complexity of a non collocated site and the use of 30 seconds samples. In [Pedro et al. \(2018\)](#) a FS of 8.6% and 13.3% is obtained for 5 minutes ahead forecast when using features extracted from all-sky images (such as the average, standard deviation and entropy values) in two machine learning approaches, k -Nearest Neighbour and Gradient Boosting, respectively. These similar results are obtained over a one year data set which include all sky condition days and not only highly variable ones. A quick comparison with pure time series models can be done by mentioning the results obtained in [Paulescu et al., 2014](#), in which an ARIMA model is used to predict GHI 5 minutes ahead. The obtained relative RMSD for the whole data set (five consecutive days in July, 2010) is about 30%, although this data set includes cloudy and sunny days. At the former, comparable to our data set, the relative RMSD is as high as 40%. For a fine tune comparison, information of the cloud type and scenes are required. Finally, in terms of RDI, we found similar results that the ones presented in [Chu et al., 2015](#), where a RDI of 67.3% is reported for high-magnitude ramps and 55.0% for moderate ramps, both for a forecast horizon of 10 minutes. In a general comparison with [Chu et al. \(2015\)](#), the proposed technique achieves higher forecasting skills but lower ramp detection indexes.

To include further details of the performance evaluation, Table 5 shows the highest FS and RDI in each day for the partly cloudy data set, along with the corresponding one-minute variability of clear-sky index, σ_c . The first thing to note is that there appears to be no direct correspondence between variability and performance. Rather, it appears that the type of the all-sky scene, i.e. cloud types and their sky distribution, is the main factor affecting the performance in each day. The second conclusion obtained from this Table (jointly with Table 1) is that highest FS with respect to regular persistence at longer forecast horizons ($\Delta t_i > 5$ minutes) is achieved when cumulus clouds type prevail, as on 21st May, 21st June and 29th November. This is due to the fact that this type of clouds are not only easily detected by the cloud detection algorithm, but also the CCM cloud displacement algorithm gives more accurate results when dealing with one layer of this kind of clouds (on the 21st May multiple cloud layer situations do occur, but the prevailing situation is that of one layer of cumulus cloud type). However, not only cloud types and number of layers affect the performance of the proposed model, but also the position of clouds relative to the sun, and the direction of the cloud motion. This explains the relatively low performances achieved on days which also presented mostly cumulus type clouds (7th and 13th September). In the remaining days, multiple cloud layers often coexisted in a same image, leading to inaccurate results for longer forecast horizons, since as stated before this situation poses a challenge for the CCM cloud displacement algorithm. In such cases, the information extracted from the mean displacement vector (grid orientation and cell size, cloud fractions X_i , and predicted k_c) represents the motion of one particular cloud layer, as explained in 3.3, and the effect of the remaining layers is not considered, which leads to a worse performance as compared to a single cloud layer.

As mainly the cloud types and their distribution in the sky dome affect the performance of the system, let us analyze also a day in particular, the 26th November, where different types of low (Cu, Sc) and high (Cs, Ac) clouds coexisted (see Table 1).

Figure 2, but the 1-minute variability.

Date	σ_c	Highest FS		Highest RDI	
		regular	smart	HMR	MMR
17/05/2016	0.17	5.2 (10)	9.2 (1)	75.0 (1)	45.8 (4)
21/05/2016	0.19	16.9 (9)	14.9 (2)	80.0 (1)	62.5 (4)
02/06/2016	0.16	15.9 (4)	10.6 (1)	80.0 (2)	57.1 (3)
05/06/2016	0.19	12.8 (5)	21.1 (2)	63.6 (3)	53.3 (10)
21/06/2016	0.20	19.6 (8)	2.7 (1)	75.0 (1)	73.3 (8)
06/07/2016	0.20	3.5 (7)	-2.7 (8)	43.8 (7)	52.9 (8)
15/07/2016	0.19	4.7 (8)	-4.6 (1)	84.6 (1)	50.0 (8)
18/07/2016	0.15	7.6 (10)	2.6 (9)	66.7 (1)	55.6 (2)
07/09/2016	0.24	18.6 (5)	10.1 (1)	78.3 (1)	60.0 (4)
13/09/2016	0.42	13.3 (10)	-0.7 (1)	67.2 (1)	78.6 (8)
29/09/2016	0.14	12.9 (10)	8.6 (2)	57.1 (2)	90.9 (1)
30/09/2016	0.26	17.1 (4)	6.7 (3)	73.3 (5)	60.6 (8)
06/10/2016	0.24	21.7 (6)	20.8 (1)	78.8 (9)	63.2 (6)
07/10/2016	0.15	3.0 (10)	7.2 (10)	66.7 (2)	50.0 (1)
26/11/2016	0.18	22.8 (4)	15.7 (4)	70.3 (4)	52.0 (3)
29/11/2016	0.29	21.9 (8)	11.2 (8)	60.0 (3)	77.8 (6)

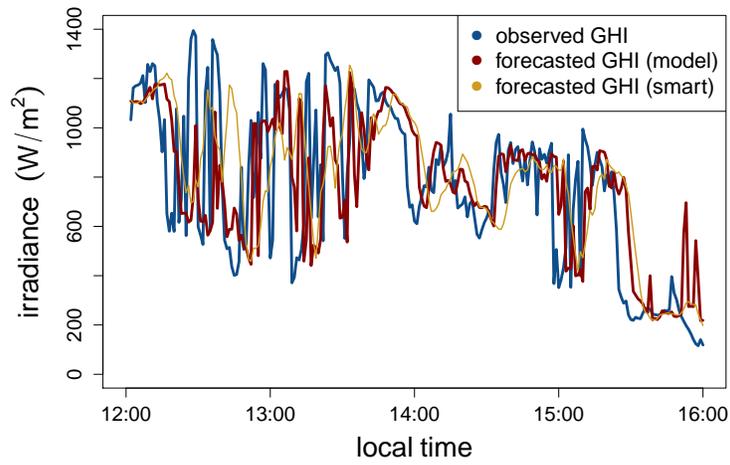
Table 5: Highest performance values for each day in the partly cloudy sub data set. Numbers in parenthesis indicate corresponding forecast horizon in minutes. The 1-minute variability, σ_c , is shown in the second column. Both FS and RDI are expressed as a percentage.

As shown in Table 5, this day also show the highest FS with respect to regular persistence of all the partly cloudy days but not with respect to the smart persistence. In addition, relatively low RDI are obtained for this day. Figure 12-(a) and (b) show respectively, for this day, the intermittent GHI and k_c time-series (in blue), and their 5 minutes ahead forecast, by the model (in red) and by the smart persistence (in yellow). Figure 12-(c) shows the average cloud fraction time-series for this day, which varies between 17% and 99% with a mean value of 64%. The absolute RMSD for the 5 minutes ahead model’s forecast for this day was 253 W/m² (rRMSD \approx 33%), while the persistences were 313 W/m² (regular) and 283 W/m² (smart), yielding a FS of 19.3% and 10.7% in comparison to each persistence procedure, respectively. Regarding ramp detection, the model achieved –for high-magnitude ramps (HMR)– a maximum RDI (70%) at $\Delta t_i = 4$ minutes, and a minimum RDI of more than 36% at $\Delta t_i = 1$ minute. For moderate-magnitude ramps (MMR) the resulting RDI peaked at $\Delta t_i = 3$ minutes (RDI = 52%) and had a minimum at $\Delta t_i = 10$ minutes (RDI = 31%). As shown in Figure 13, about 26% of the samples correspond to optically thick clouds covering the sun ($k_c < 0.6$), 31% represent optically thin clouds covering the sun ($0.6 < k_c < 1.0$), and the remaining 43% represent clear sun disk with nearby clouds. A part of this sample (about 10%) contains also enhancement events, represented by clear-sky indexes above 1.2.

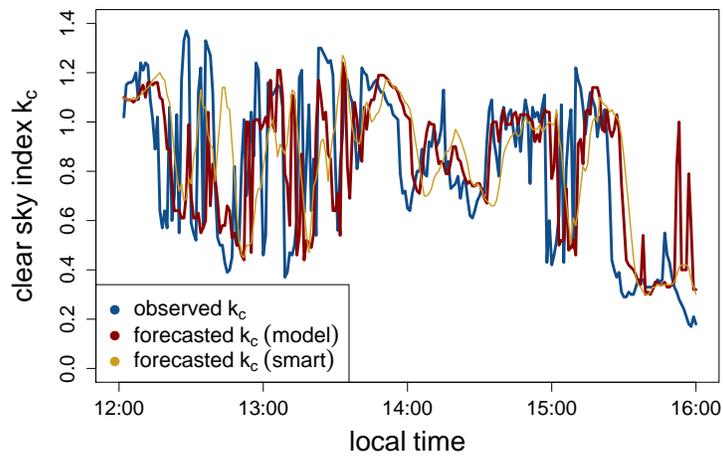
If we consider the whole data set of Table 1 (partly cloudy, overcast and clear sky), the model also starts outperforming persistence at a forecast horizon of 2 minutes (FS = 3.3%), and its performance peaks also at 6 minutes but with a rather lower forecasting skill (FS = 9.9%), due to the influence of overcast sky, as will be discussed in next Subsection.

4.3. Results for clear and overcast days

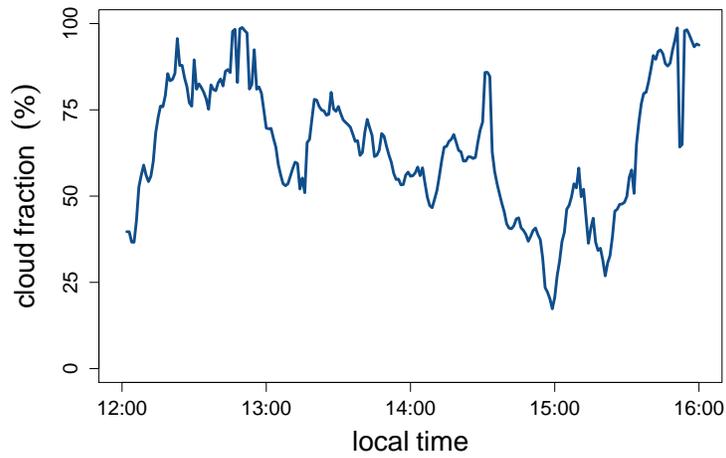
The proposed model is intended to provide GHI forecasts for partly cloudy conditions where the occurrence of ramps may affect the PV power output. For the sake of completeness, its performance on two other separate sets of data is also evaluated, one composed only by clear-sky days and the other only by overcast days. The clear-sky and overcast data sets contains 698 and 718 forecast time points, respectively, and the dates are provided in Table 1. For clear-sky days, the proposed model outperforms both persistence models at all forecast horizons. FS with respect to regular persistence reaches a maximum of 28.2% at $\Delta t_i = 4$ minutes. FS with respect to smart persistence decreases practically monotonically with forecast horizon,



(a) Measured and 5-minutes ahead forecast for G_H .



(b) Measured and 5-minutes ahead forecast for k_c .



(c) Cloud fraction time-series.

Figure 12: Time-series for the 26th November, 2016, between 12:00 and 16:00 local time.

446 starting with a maximum of 21.7% (1 minute ahead) down to a minimum of 16.3% (10 minutes ahead). Despite these high
 447 levels of FS, RMSD levels for the three methods are below $6 W/m^2$, and the absolute difference between them is small, as

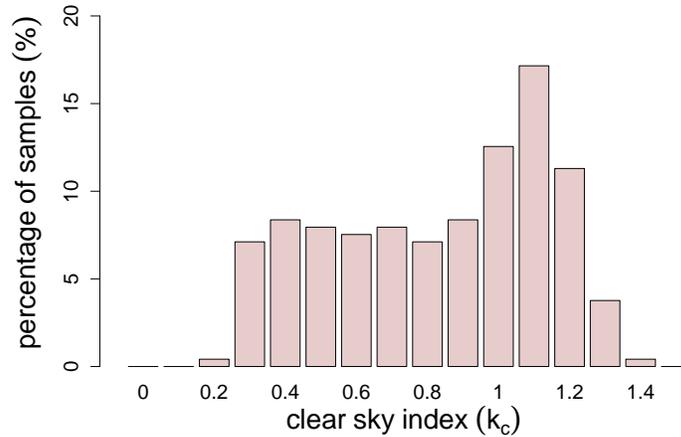


Figure 13: Clear-sky index distribution for the 26th November, 2016, between 12:00 and 16:00 local time.

448 shown in Figure 14-(a). For overcast days, the proposed model does not outperform persistence for any forecast horizon.
 449 RMSD of the proposed model for overcast days ranges from 64 W/m² to 110 W/m², while both persistence models have
 450 a RMSD between 25 W/m² and 45 W/m². The relatively high RMSD of the proposed model for overcast days is due to
 451 misclassification of cloudy pixels as clear when the sky is dark enough so that the global RBR threshold used (0.9) is not low
 452 enough (as explained in Subsection 3.2). This misclassification introduces spurious spikes in the predicted k_c (and GHI) which
 453 raise the RMSD metric. Although this misclassification can be mitigated by the use of an adaptive global RBR threshold, we
 454 chose not to modify the use of a fixed threshold of 0.9 since it yields optimal cloud detection results under partly cloudy
 455 conditions, which is the focus of the present work.

456 The behavior of the proposed and persistence models for the three sub data sets presented here (partly cloudy, overcast,
 457 clear sky) suggests the use of a hybrid model that uses either a persistence model (regular or smart) or the proposed model
 458 depending on sky conditions and forecast time horizon.

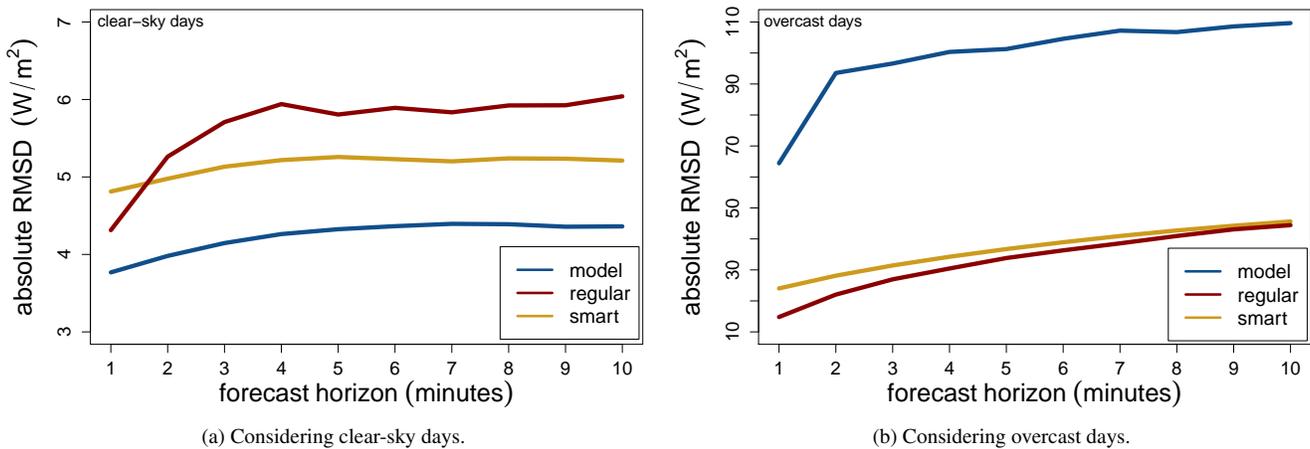


Figure 14: Absolute root mean squared deviation for the persistence and the modified persistence as a function of the forecast horizon. The figures illustrate the proposed model's different performance for overcast and clear-sky days.

459 5. Conclusions

460 In this paper, a method for very short-term solar forecasting which combined the use of all-sky images and real-time solar
 461 irradiance time series was proposed and evaluated for a testing site in the south of South America. The images were acquired

Δt_i	clear sky condition			overcast sky condition		
	model	regular	smart	model	regular	smart
1 min	3.8	4.2	4.5	64	15	24
2 min	4.0	5.0	4.7	94	22	28
3 min	4.2	5.4	4.9	97	27	31
4 min	4.3	5.6	5.0	100	30	34
5 min	4.3	5.6	5.0	101	34	37
6 min	4.4	5.6	5.0	105	36	39
7 min	4.4	5.7	5.0	107	38	41
8 min	4.4	5.7	5.0	107	41	43
9 min	4.4	5.7	5.0	109	43	44
10 min	4.4	5.8	4.9	110	44	46

Table 6: Absolute root mean squared deviation (in W/m^2) for the proposed model, regular persistence and smart persistence for overcast and clear-sky days.

with a locally designed low-cost all-sky imager, based on an off-the-shelf camera and lens. The purpose of the system is to predict one-minute averaged GHI, as measured by a point sensor, with the use of an all-sky imager collocated with the sensor. The all-sky imager acquired one image per minute. These images were used to estimate the mean motion of clouds using pre-existing all-sky image processing techniques with some adaptations and with the addition of a simple post-processing procedure. This mean cloud motion was used to generate a ladder-formed grid starting from the sun and oriented in the opposite direction to the mean cloud displacement vector, a method known in the literature as “grid cloud-fraction method”. The impact of clouds covering the sun in the solar irradiance was quantified with the use of measured GHI time-series. A simple deterministic model (Eq. (6)) is proposed to forecast the clear-sky index (and GHI) from 1 to 10 minutes ahead, based on the grid cloud fraction and real-time measurements information. The results show that in terms of forecasting skill (relative to regular persistence) the proposed method achieved a peak performance of 11.4% at forecast horizons of 6 and 10 minutes under partly cloudy conditions. In such cases, the model started outperforming regular persistence at a forecast horizon of 2 minutes. When compared to a smart persistence model based on the past 5 minutes k_c average (the same time-window length used for real-time processing), the proposed model outperformed it at all forecast time horizons, although moderately (FS between 2 and 3%, approximately). The proposed method’s ability to forecast ramps was also evaluated by calculating the ramp detection index (RDI) as a function of forecast horizon. Two types of ramps were considered, namely high-magnitude ramps and moderate ramps. For high-magnitude ramps RDI had small variations with forecast horizon, ranging from 55% to 62%, approximately. For moderate ramps, RDI showed also smooth variations with forecast horizon, with a peak performance at 8 minutes equal to about 49%. These results show that the proposed methodology is suitable to forecast short term GHI values and ramps for the challenging highly variable solar irradiance conditions, like partly cloudy. The model also exhibited good performance under clear-sky conditions, outperforming both persistence models at all forecast horizons. On the other hand, for overcast days the model did not outperform any of the persistence models because spurious clear-sky pixels were misclassified in some overcast sky images. Although this behavior could be improved by the use of an adaptive RBR threshold, we prioritize the use of a fixed RBR threshold as it is more accurate to detect clouds under partly cloudy conditions, the focus of this work. Thus, one of the main advantages of the proposed model is its ability to forecast solar irradiance under highly variable conditions and challenging cloud conditions, through a simple formulation where only two easily calculated parameters are extracted from real time irradiance measurements.

Further improvements can be done in the proposed system. Thin cloud detection and processing of the circumsolar region could be enhanced by the use of state-of-the-art HDR imaging systems together with improvements in the cloud detection algorithm. This will lead to fewer deviations under partly cloudy and mostly clear-sky conditions. Also, an adaptive RBR threshold may be used to improve the performance under overcast sky conditions. For the estimate of the mean cloud displacement, the number of ladder cells (i.e forecast horizons) could be adaptive, allowing for larger forecasts when possible (i.e depending on cloud speed and direction). Also, robust optical flow techniques (Black and Anandan, 1996) or other variational approaches (Horn and Schunck, 1981), that are well-known in the field of computer vision, may be an interesting

495 alternative to derive the cloud motion fields from the images. This would allow a dense estimation of smoother velocity
496 fields and may reduce filtering stages that are required to obtain the mean cloud displacement from two consecutive images.
497 Finally, a hybrid real-time solar forecasting system for the very short term, based on the proposed model and persistence
498 could be designed, by making use of a regular and/or a smart persistence model for the situations in which it achieves better
499 performance, namely for $\Delta t_i < 2$ minutes in partly cloudy, and for all Δt_i in overcast sky conditions. This might be achieved
500 by integrating these forecasts with machine learning techniques. Some of these proposals are part of our current work.

501 Acknowledgments

502 The authors gratefully acknowledge the partial financial support given by the National Research and Innovation Agency
503 (ANII, Uruguay) through the grant FSE-ANII-1-2013-1-10613, and by the National Electricity Market Administration (ADME,
504 Uruguay) and the Corporación Andina de Fomento (CAF) through the project PRONOS. We also wish to thank Dr. Gonzalo
505 Abal and Dr. Gonzalo Tancredi for their academic support, and Javier Capeche for the preventive and corrective maintenance
506 of the hardware.

507 References

- 508 Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., de Pison, F. M., and Antonanzas-Torres, F. (2016). Review of photovoltaic
509 power forecasting. *Solar Energy*, 136:78–111.
- 510 Black, M. J. and Anandan, P. (1996). The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields.
511 *Computer Vision and Image Understanding*, 63(1):75–104.
- 512 Chauvin, R., Nou, J., Thil, S., and Grieu, S. (2015). Modeling the clear-sky intensity distribution using a sky imager. *Solar*
513 *Energy*, 119:1–17.
- 514 Chow, C. W., Belongie, S., and Kleissl, J. (2015). Cloud motion and stability estimation for intra-hour solar forecasting. *Solar*
515 *Energy*, 115:645–655.
- 516 Chow, C. W., Urquhart, B., Lave, M., Dominguez, A., Kleissl, J., Shields, J., and Washom, B. (2011). Intra-hour forecasting
517 with a total sky imager at the UC San Diego solar energy testbed. *Solar Energy*, 85(11):2881–2893.
- 518 Chu, Y., Pedro, H. T., and Coimbra, C. F. (2013). Hybrid intra-hour DNI forecasts with sky image processing enhanced by
519 stochastic learning. *Solar Energy*, 98:592–603.
- 520 Chu, Y., Pedro, H. T., Li, M., and Coimbra, C. F. (2015). Real-time forecasting of solar irradiance ramps with smart image
521 processing. *Solar Energy*, 114:91–104.
- 522 Coimbra, C. F., Kleissl, J., and Marquez, R. (2013). Chapter 8 - Overview of solar-forecasting methods and a metric for
523 accuracy evaluation. In Kleissl, J., editor, *Solar Energy Forecasting and Resource Assessment*, pages 171–194. Academic
524 Press, Boston.
- 525 de la Parra, I., J.Marcos, M.García, and L.Marroyo (2015). Storage requirements for pv power ramp-rate control in a PV fleet.
526 *Solar Energy*, 118:426–440.
- 527 Diagne, M., David, M., Lauret, P., Boland, J., and Schmutz, N. (2013). Review of solar irradiance forecasting methods and a
528 proposition for small-scale insular grids. *Energy Reviews*, 27:65–76.
- 529 Ela, E., Diakov, V., Ibanez, E., and Heaney, M. (2013). Impacts of variability and uncertainty in solar photovoltaic generation
530 at multiple timescales. *National Renewable Energy Laboratory (NREL) Technical report*, NREL/TP-5500-58274.
- 531 Fu, C.-L. and Cheng, H.-Y. (2013). Predicting solar irradiance with all-sky image features via regression. *Solar Energy*,
532 97:537–550.
- 533 Ghonima, M. S., Urquhart, B., Chow, C. W., Shields, J. E., Cazorla, A., and Kleissl, J. (2012). A method for cloud detection
534 and opacity classification based on ground based sky imagery. *Atmospheric Measurement Techniques*, 5(11):2881–2892.

- 535 Giacosa, G. and Alonso-Suárez, R. (2018). Desempeño de la persistencia para la predicción del recurso solar en Uruguay. In
536 *VII Brazilian Solar Energy Congress (CBENS)*.
- 537 Hamill, T. M. and Nehrkorn, T. (1993). A short-term cloud forecast scheme using cross correlations. *Weather and Forecasting*,
538 8(4):401–411.
- 539 Hoff, T. E. and Perez, R. (2010). Quantifying PV power output variability. *Solar Energy*, 84(10):1782–1793.
- 540 Horn, B. K. and Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence*, 17(1):185–203.
- 541 Ineichen, P. and Perez, R. (2002). A new airmass independent formulation for the Linke turbidity coefficient. *Solar Energy*,
542 73(3):151–157.
- 543 Kasten, F. (1980). A simple parameterization of the pyrheliometric formula for determining the Linke turbidity factor. *Meteor.*
544 *Rundschau*, 33:124–127.
- 545 Kleissl, J. (2013). *Solar Energy Forecasting and Resource Assessment*. Academic Press, Oxford.
- 546 Laguarda, A. and Abal, G. (2017). Clear-sky broadband irradiance: First model assessment in uruguay. In *ISES Solar World*
547 *Congress 2017*.
- 548 Lauret, P., Perez, R., Aguiar, L. M., Tapachès, E., Diagne, H. M., and David, M. (2016). Characterization of the intraday
549 variability regime of solar irradiation of climatically distinct locations. *Solar Energy*, 125(Supplement C):99–110.
- 550 Lave, M., Kleissl, J., and Arias-Castro, E. (2012). High-frequency irradiance fluctuations and geographic smoothing. *Solar*
551 *Energy*, 86(8):2190–2199. Progress in Solar Energy 3.
- 552 Lave, M., Kleissl, J., and Stein, J. (2013). Chapter 7 - quantifying and simulating solar-plant variability using irradiance data.
553 In Kleissl, J., editor, *Solar Energy Forecasting and Resource Assessment*, pages 149–169. Academic Press, Boston.
- 554 Lefèvre, M., Oumbe, A., Blanc, P., Espinar, B., Gschwind, B., Qu, Z., Wald, L., Schroedter-Homscheidt, M., Hoyer-Klick, C.,
555 Arola, A., Benedetti, A., Kaiser, J. W., and Morcrette, J.-J. (2013). McClear: a new model estimating downwelling solar
556 radiation at ground level in clear-sky conditions. *Atmospheric Measurement Techniques*, 6(9):2403–2418.
- 557 Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied*
558 *Mathematics*, 2:164–168.
- 559 Li, Q., Lu, W., and Yang, J. (2011). A hybrid thresholding algorithm for cloud detection on ground-based color images.
560 *Journal of atmospheric and oceanic technology*, 28:1286–1296.
- 561 Marcos, J., Marroyo, L., Lorenzo, E., and García, M. (2011). Smoothing of PV power fluctuations by geographical dispersion.
562 *Prog. Photovolt: Res. Appl.*, 20:226–237.
- 563 Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathe-*
564 *matics*, 11:431–441.
- 565 Marquez, R. and Coimbra, C. F. (2013). Intra-hour DNI forecasting based on cloud tracking image analysis. *Solar Energy*,
566 91:327–336.
- 567 Nou, J., Chauvin, R., Eynard, J., Thil, S., and Grieu, S. (2018). Towards the intrahour forecasting of direct normal irradiance
568 using sky-imaging data. *Heliyon*, 4(4):e00598.
- 569 Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cyber-*
570 *netics*, 9(1):62–66.
- 571 Paulescu, M., Mares, O., Paulescu, E., Stefu, N., Pacurar, A., Calinoiu, D., Gravila, P., Pop, N., and Boata, R. (2014).
572 Nowcasting solar irradiance using the sunshine number. *Energy Conversion and Management*, 79:690–697.
- 573 Pedro, H. T., Coimbra, C. F., David, M., and Lauret, P. (2018). Assessment of machine learning techniques for deterministic
574 and probabilistic intra-hour solar forecasts. *Renewable Energy*, 123:191–203.

- 575 Quesada-Ruiz, S., Chu, Y., Tovar-Pescador, J., Pedro, H., and Coimbra, C. (2014). Cloud-tracking methodology for intra-hour
576 DNI forecasting. *Solar Energy*, 102:267–275.
- 577 Remund, J., Wald, L., Lefevre, M., Ranchin, T., and Page, J. (2003). Worldwide Linke turbidity information. *Proceedings of*
578 *ISES Solar World Congress*.
- 579 REN21 (2018). Renewables 2018 global status report. Technical Report 978-3-9818107-6-9, Renewable Energy Policy
580 Network for the 21st. Century, Paris: REN21 Secretariat.
- 581 Rigollier, C., Bauer, O., and Wald, L. (2000). On the clear sky model of the ESRA - European Solar Radiation Atlas - with
582 respect to the Heliosat method. *Solar Energy*, 68(1):33–48.
- 583 Rowlands, I. H., Kemery, B. P., and Beausoleil-Morrison, I. (2014). Managing solar PV variability with geographical disper-
584 sion: An Ontario (Canada) case-study. *Renewable energy*, 68:171–180.
- 585 Scaramuzza, D., Martinelli, A., and Siegwart, R. (2006). A toolbox for easily calibrating omnidirectional cameras. In *2006*
586 *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5695–5701.
- 587 Senjyu, T., Datta, M., Yona, A., Funabashi, T., and Kim, C.-H. (2008). PV output power fluctuations smoothing and optimum
588 capacity of energy storage system for PV power generator. *Renewable Energy and Power Quality Journal*, 1:213.
- 589 Urquhart, B., Kurtz, B., Dahlin, E., Ghonima, M., Shields, J. E., and Kleissl, J. (2015). Development of a sky imaging system
590 for short-term solar power forecasting. *Atmospheric Measurement Techniques*, 8:875–890.
- 591 Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F., and Fouilloy, A. (2017). Machine learning methods
592 for solar radiation forecasting: A review. *Renewable Energy*, 105:569–582.
- 593 Yang, D., Yagli, G. M., and Quan, H. (2018). Quality control for solar irradiance data. *2018 IEEE Innovative Smart Grid*
594 *Technologies - Asia (ISGT Asia)*, pages 208–213.
- 595 Yang, H., Kurtz, B., Nguyen, D., Urquhart, B., Chow, C. W., Ghonima, M., and Kleissl, J. (2014). Solar irradiance forecasting
596 using a ground-based sky imager developed at UC San Diego. *Solar Energy*, 103:502–524.