

Intra-hour DNI forecasting based on cloud tracking image analysis

Ricardo Marquez^a, Carlos F.M. Coimbra^{b,*}

^a *Mechanical Engineering and Applied Mechanics Graduate Program, School of Engineering, University of California, Merced, CA 95343, USA*

^b *Department of Mechanical and Aerospace Engineering, Jacobs School of Engineering, Center for Renewable Resource Integration, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA*

Received 3 January 2012; received in revised form 11 September 2012; accepted 29 September 2012

Available online 31 October 2012

Communicated by: Associate Editor David Renne

Abstract

We describe an image processing methodology using Total Sky Imagers (TSIs) to generate short-term forecasts of Direct Normal Irradiance (DNI) at the ground level. Specifically, we are interested in forecasting 1-min averaged DNI values for time horizons varying from 3 to 15 min. This work describes several sky image processing techniques relevant to solar forecasting, including velocity field calculations, spatial transformation of the images, and cloud classification. The solar forecasts derived here are analyzed and quantified in terms of root mean square error (RMSE) deviations in relation to actual values and compared to the performance of persistence models. © 2012 Elsevier Ltd. All rights reserved.

Keywords: Solar forecasting; Sky imaging; Cloud classification; Beam radiation; DNI prediction

1. Introduction

The variability of the solar energy resource at the ground level makes it difficult to manage and integrate power output from solar farms to the power grid. Today it is widely acknowledged by power producers, utility companies and independent system operators (ISOs) that there is a need for enhanced forecasting of weather systems and their effect on solar power output. For large scale solar applications, the amount of irradiance reaching the ground surface is most strongly affected by clouds, aerosols and water vapor column, with cloud effects being particularly dominant for most locations where solar power is currently produced. Sky imagers, such as the Total Sky Imager (TSI) by Yankee Environmental Systems, originally developed for meteorological and aviation purposes, are potentially helpful for developing short-term solar irradiance prediction and forecasting models, although the time horizon covered by such instruments is limited by the relative low

image resolution and limited field of view. Here we propose and test new cloud-tracking methods for predicting future values of Direct Normal Irradiance (DNI) using processed images taken by a ground-based TSI.

Different solar irradiance forecast methodologies have been proposed for various time horizons. For the forecasting horizon of same-day to a few days ahead, NWP models have been evaluated in various locations (Perez et al., 2010; Lorenz et al., 2009; Marquez and Coimbra, 2011; Mathiesen and Kleissl, 2011; Changsong et al., 2011). For time horizons of 1–6 h, satellite-based forecasts are competitive with NWP models (Perez et al., 2010). Satellite models have also been applied to 30-min ahead forecasting (Hammer et al., 1999). In addition to these physical model/image processing-based models, time series-based solar forecasting methods have also been extensively applied, (e.g. Mellit, 2008; Mellit and Pavan, 2010; Mellit et al., 2010; Martin et al., 2010; Cao and Lin, 2008; Sfetsos and Coonick, 2000; Marquez et al., 2013; Marquez and Coimbra, 2013) and validated over wide ranges of time horizons. In some of these time-series applications, direct cloud information is not used (Sfetsos and Coonick,

* Corresponding author. Tel.: +1 858 534 4285; fax: +1 858 534 7078.
E-mail address: cocoimbra@ucsd.edu (C.F.M. Coimbra).

2000; Martin et al., 2010; Mellit and Pavan, 2010; Mellit et al., 2010; Marquez and Coimbra, 2013) while some incorporate cloud information as exogenous inputs (Cao and Lin, 2008; Marquez et al., 2013). The time-series forecasting methodologies have been applied successfully for 5–15 min (Reikard, 2009), 1-h (Cao and Lin, 2008; Sfetsos and Coonick, 2000; Martin et al., 2010), and 24-h ahead (Mellit et al., 2010).

The use of sky imagers for solar irradiance forecasting has been limited. One study (Crispin et al., 2008) employed a TSI to extract cloud features using a Radial Basis Function Neural Network (RBFNN) pixel classifier to include in an Artificial Neural Network (ANN) forecast model for 1–60-min ahead horizons. Results from this study were mixed, showing only marginal improvements with reference to a model which uses only irradiance lagged values as input variables. This initial study concluded that errors introduced by the cloud index estimates limited the accuracy of the TSI-based solar irradiance forecasts. The work of Marquez et al. (2013) uses cloud indices obtained from a TSI built-in cloud classification algorithms, and cloud indices derived from Infrared Radiometric (IR) measurements to report improved results for 1-h forecasting of Global Horizontal Irradiance (GHI) using ANNs for time-series predictions. More recently, Chow and coworkers (Chow et al., 2011) used TSI images to produce nowcasting of GHI single-point solar irradiance sensors located at the University of California San Diego meteorological network. Their study also analyzed the cloud field propagation by computing matching errors of the future image given current images translated in the direction of the computed velocity field. Their results suggest that TSI maybe useful for forecasting horizons of up to 15–20 min, depending on cloud velocities and field of view.

In the present work, detailed image processing methods aimed at forecasting DNI are presented. Our model is only tested for a single point sensor. So, in order to apply the methodology to a large solar power plant, our model assumes that the sky imager is co-located with the plant. The applicability of the methodology will depend on the

size of the plant with larger plants perhaps requiring multiple imagers and/or a cloud shadow based solar forecasting methodology such as the one presented in Chow et al. (2011).

Section 2 briefly describes the data collection procedure. Sections 3–5 describe processing methods for the images obtained by our TSI. The processing includes the determination of masks, spatial transformations, the computation of the velocity fields, and classification of cloud and clear pixels. These processing steps are exemplified in Fig. 1. In Section 6, the methodology for computing grid cloud fractions is described. This methodology defines the variables used for generating the solar forecasts. Solar forecasting errors are computed and evaluated in Section 7 with respect to the Root-Means-Squared Errors (RMSEs), and in relation to a persistent model, which is used as the baseline for comparisons with other solar forecasting methods in the literature. Conclusions follow in Section 8.

2. Data

2.1. Total Sky Imager

Images of the sky are taken at 1 min intervals at the University of California Merced's solar observatory station using Yankee Environmental Systems' TSI-880 instrument (YES, 2011). The TSI is composed of a heated rotating hemispherical mirror with a down-pointing CCD camera located above it. The mirror contains a sun tracking shadow-band that continuously covers the mirror from the direct sunlight in order to protect the camera sensor from the sun's reflection. The TSI built-in algorithms do not produce cloud classification and dynamic properties such as cloud type, speed, or height, but does produce processed images with pixels representing clear, thin-clouds, and opaque-clouds. The built-in cloud classification algorithm relies on simple red to blue pixel intensity thresholds, which were developed for cloud fraction estimations. For the forecasting application under consideration here, we instead use our own algorithms (described in Section 5),

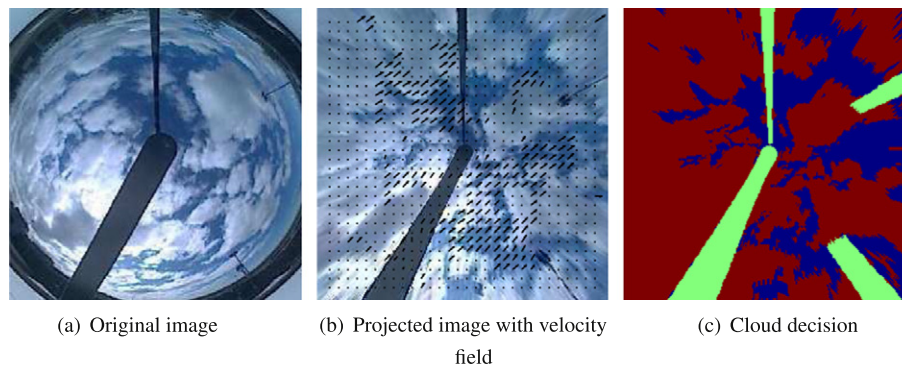


Fig. 1. Main image processing steps. *Left*: Original 8-bit image in RGB color space. *Middle*: Image projected to rectangular grid using derived image to sky mapping, and velocity fields computed using piv algorithm. *Right*: Cloud decision image based on MCE algorithm. Green areas represent false image pixels which are neglected in cloud decision algorithm. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

since the built-in cloud classifier does not produce accurate classifications in areas close to the sun's location in the images.

2.2. Solar irradiance data

High quality solar irradiance data has been collected at the UC Merced Solar Observatory since September 2009. The solar instrumentation used in this study includes Eppley Precision Spectral Pyranometers (PSPs) and Normal Incidence Pyrheliometers (NIPs), with one shaded PSP and one NIP mounted on an Eppley model SMT tracker. The TSI is located within 5 m of the rest of the solar instruments. Data is logged with a Campbell Scientific CR1000 datalogger at a sampling rate of 2 s from which 30-s averages are recorded. In this work, we included only DNI data as measured by an Eppley NIP, which typically deviates from true Absolute Cavity Radiometer (ACR) values by less than 3% (not including obvious errors due to misalignments caused by tracker stoppage). When broken clouds are present, instrumentation errors of 3% are very small compared to the frequent fluctuations in DNI that can account for full value drops (100%), and amplifications due to cloud edge effects. Predicting these fluctuations is the main goal of this work.

3. Processing sky images

Relevant image processing procedures are presented in this section for short-term solar forecasting applications. First, suitable masks are created for the images so that segments such as the protective shadow band and some nearby object appearing on the horizon (such as poles and fences that are not part of the sky) can be discarded. Next, we transform the images onto a rectangular grid so that clouds and velocity fields are more accurately represented.

3.1. Masks

The shadow band's location is a critical obstruction for DNI predictions, and severely limits the ability of using the TSI for very-short term solar forecasting. Additionally, the glare in the circumsolar region makes it difficult to apply easy rules for replacing the shadow-band pixels with an accurate representation of the sky. These difficulties are avoided here by focusing on forecasts with horizons larger than 3 min, and by applying masks to effectively ignore the effects of areas close to the solar region. The masks are matrices with binary elements corresponding to locations of the image. The element values are used to indicate true or false pixel values. Regions indicated as true are the regions of interest (ROI) and only these pixels are used in the cloud decision and forecasting exercises.

The binary masks are created by making use of predictable attributes of the images:

1. The images are always enclosed by a circle located of the same size and centered at a constant *position*.
2. The shadow-band and nearby objects in the horizons such as poles are always the same size and shape.
3. The shadow-band moves in a predictable fashion, according to the solar azimuth angle.

The task of determining the circumference of the ROI is a suitable application of the function *CircularHough_Grd* available at the Matlab Central File Exchange (Peng, 2011). This program was developed for detecting various radii in grayscale images via Hough Transforms (Peng et al., 2007) and is used here to detect the outer edges of the sky imager dome, and then to compute a best fit circle by estimating the proper radius and circle center. The orientation of the shadow-band is determined using solar azimuth calculations obtained from Duffie and Beckman (2006). An example of an ROI is the white regions shown in Fig. 2.

3.2. Transformations to rectangular grid

This section covers the approach used to transform the images from a spherical to a rectangular grid. The following approach is based on knowledge of the solar zenith angles and the sun's apparent position in the image to arrive at the desired transformation. The image pixel locations are first represented in polar coordinates,

$$x_I = r_I \cos(\phi), \quad y_I = r_I \sin(\phi), \quad (1)$$

where r_I is the radial distance from the center of the TSI raw image to the coordinate $[x_I, y_I]$, and ϕ is the azimuth angle with respect to the horizontal axis and refer to the image I . The rectangular grid projected on the sky (S) coordinates are similarly represented as,

$$x_S = r_S \cos(\phi), \quad y_S = r_S \sin(\phi), \quad (2)$$

where r_S is the radial distance for rectangular grid coordinates on the sky. Corresponding pixels in either the I or S coordinate frame share the same values for the polar angle

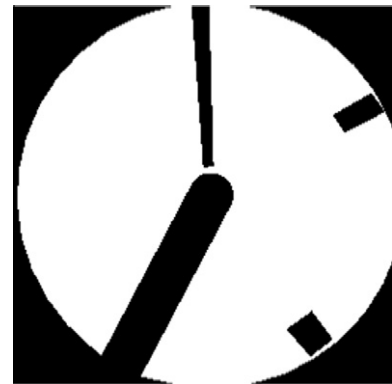


Fig. 2. Example of an ROI (region-of-interest) as represented by the white areas. Note that the TSI shadow band and nearby objects in the horizon are not included in the ROI.

ϕ , therefore, the mapping is accomplished simply by determining the relationship between r_I and r_S . This relationship is obtained by considering the sky zenith angle (θ). The sky radial distance r_S can be computed given the zenith angle as,

$$r_S = H \tan^{-1}(\theta). \quad (3)$$

where H is a reference height of the sky, such as the cloud base height. For any given time, we can determine the solar zenith angle (θ_{SZA}), and the sun's position on the image ($r_{I,\text{sun}}$), again, using the Hough transform algorithm (Peng, 2011). The mapping $r_S \rightarrow r_I$ is approximated by curve fitting data of r_I for the sun with θ_{SZA} then applying Eq. (3). We selected three days and computed θ_{SZA} and $r_{I,\text{sun}}$, and determined a best fit cubic polynomial is adequate for the mapping, since the y -intercept is close to zero and the curvature of the data tendency were accurately modeled (see Fig. 3).

In summary, the mapping from the sky position to the image position is,

$$r_S = \sqrt{x_S^2 + y_S^2} \quad (4)$$

$$\theta = \tan^{-1}(r_S/H) \quad (5)$$

$$r_I = r_I(\theta; \text{according to curve fit results}) \quad (6)$$

$$[x_I, y_I] = r_I[\cos(\phi), \sin(\phi)]. \quad (7)$$

The above set of equations define the projections from the apparent positions in the image to a flat sky. In the present forecasting application, we will not require the exact position of clouds since we are not computing cloud shadows over a region as in Chow et al. (2011). The scale factor H in the above projections is removed in our application by considering the normalized variables x_S/H , y_S/H and r_S/H . An example of an original image and a corresponding image projection is shown in Fig. 1a and b, respectively.

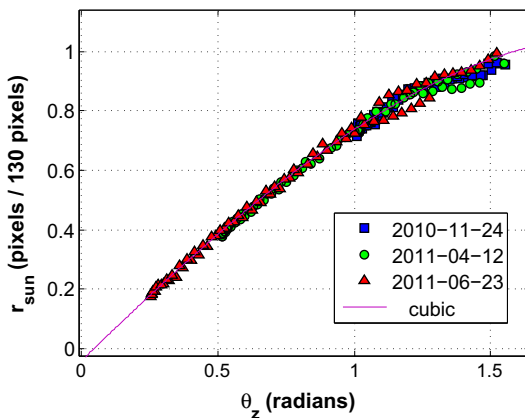


Fig. 3. Radial distance of solar position to center and solar zenith angles computed for three days as shown in legend. June 23 is solar summer solstice and is the longest day of the year. This day also has the highest solar zenith angle range of all days of the year and is therefore used for the curve-fitting.

4. Computing velocity fields

4.1. Particle image velocimetry

We use the ‘MPIV’ computer software developed by Mori and Chang (2003) to obtain the cloud velocity fields. The interrogation window is defined by a 32 pixels \times 32 pixels area, and the overlap between two consecutive windows is set to 50%, with maximum displacements set to 20 pixels for both vertical and horizontal directions. The MQD (Minimum Quadratic Difference) method (Mori and Chang, 2003) is used with two recursive checks. The velocity field is filtered and smoothed by the filter functions in the ‘MPIV’ toolbox. The sequence of images are given at 1-min apart. An example of a velocity field calculation is shown in Fig. 1b. A detailed description of the Particle Image Velocimetry (PIV) methods used here can be found in Mori and Chang (2003).

Qualitatively, the PIV algorithm seems to perform well, although it is not always possible to retrieve velocity fields from less well defined clouds. One way to evaluate the velocity vector output is to calculate matching errors, as was done by Chow et al. (2011). However, because clouds are always deforming (there are also artificial deformations due to the mirror curvature and image projections), the matching errors do not accurately represent the quality of the velocity field output. Therefore we do not compute the matching errors or further evaluate the velocity vector output other than using visual inspection. Because our ultimate goal is to develop a standalone machine learning method that learns from translations errors over time, we anticipate that some ‘learnable’ translation error can be overcome by the stochastic learning process.

4.2. Selecting a representative velocity vector

For each image, several velocity vectors are computed depending on the local motions of clouds at points (x_i, y_i) . The following analysis is simplified by selecting a representative vector for each image. A representative velocity vector is chosen by applying k -means clustering on the distribution of velocity vectors. In parts of the image where clouds are absent, velocity vectors of zero magnitude are computed and thus each image will usually have a cluster of velocity vectors near to or at the zero velocity vector. We therefore cluster each velocity field around two cluster means, which are determined using a cluster means algorithm. The larger (in magnitude) of the cluster means is chosen as the representative velocity vector. Fig. 4 shows an example of using the k -means clustering algorithm to select a single representative velocity vector.

For the example data in Fig. 4, the velocity vector: ($u = 6.0$, $v = -4.6$) pixels/min is chosen as the representative velocity vector for estimating uniform cloud motion. From this velocity, we can estimate the time that it takes for a cloud to pass through the observable range of the sky as seen by the TSI by taking the total number of pixels

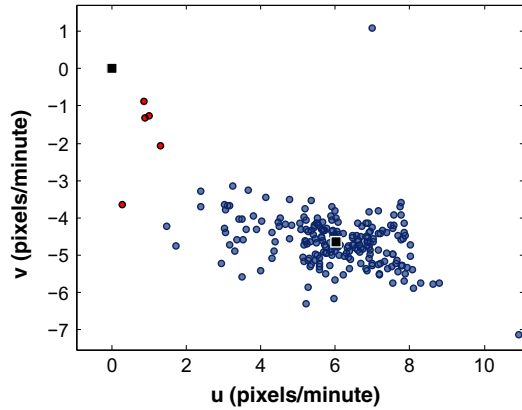


Fig. 4. Clustering velocity field using k – means algorithm. Velocity cluster mean at zero represents the clusters of velocities computed at blue parts of sky. The non-zero velocity cluster mean at $(u = 6.0, v = -4.6)$ pixels/min is selected as the representative velocity vector for the image.

(255), say in the horizontal distance, then dividing by the horizontal velocity (6.0 pixels/min), and calculating that it takes approximately 42 min for a cloud to cross the observable range of the sky for these conditions. For applications, the utility time of using the TSI for solar irradiance forecasting is about 1/3 of the estimated 42 min since more than 1 pixel is need for advection (or to be useful for making predictions). The longest time-scales for which the TSI is useful for predictions is approximately 12–18 min.

5. Cloud pixel identification

Perhaps the most important step in using a TSI-based method for solar forecasting is the identification of cloud pixels of the images. A common approach is to use the red and blue channels of the image (Long et al., 2006; Huo and Lu, 2009; Neto et al., 2010; Li et al., 2011). The clear sky is characterized by high blue pixel intensities (b) and low red pixel intensities (r), whereas, pixels with clouds are characterized by higher intensities in both channels. The distinctions are more fuzzy when there are thin clouds and bright areas caused by large amounts of scattering in the solar region of the images, making it difficult to apply a universal threshold for cloud pixel classification. Here we follow the approach recently described by Li et al. (2011), which employs an adaptive threshold scheme. The algorithm determines the cloud pixels by selecting the threshold as the minimizer of the cross-entropy between the normalized r/b ratio image and the segmented binary image (B_τ) that depends on the threshold τ .

The normalized r/b ratio defined by Li et al. (2011) is given by

$$\lambda_n = (b - r)/(b + r), \quad (8)$$

where b and r refer to the values of blue and the red channel, respectively. To identify clouds, a threshold (τ) is specified which segments the clear and cloudy parts of the sky. The new segmented image is related to the image of λ_n intensities by,

$$B_\tau(x, y) = \begin{cases} \mu(1, j) & \lambda_n(x, y) \leq \tau \\ \mu(j + 1, L) & \lambda_n(x, y) \geq \tau \end{cases} \quad (9)$$

where $\mu(a, b)$ is defined as

$$\mu(a, b) = \sum_{i=a}^b ih(i) / \sum_{i=a}^b h(i), \quad (10)$$

where $h(i)$ is the number of counts of the i th bin of the histogram of discretized levels of λ_n , denoted as $\hat{\lambda}_n$. The threshold hold is set to the corresponding minimizer of the cross-entropy between λ_n and B_τ according to

$$j^* = \arg \min \{ -m(1, j) \log(\mu(1, j)), -m(j + 1, L) \log[\mu(j + 1, L)] \},$$

where $m(a, b) = \sum_{i=a}^b ih(i)$, thus, $\tau^* = \hat{\lambda}_n(j^*)$. This classification algorithm, referred to as the MCE (Minimum Cross Entropy) algorithm, was developed and tested in Li et al. (2011) as part of the HYTA (HYbrid Threshold Algorithm). The MCE is meant to be applied to images with bimodal r/b histograms. The MCE is applicable in our case since we are mainly interested in days with mostly broken clouds where DNI is more variable.

Two modifications to the MCE algorithm were made. Firstly, we decided that whenever $\tau^* > 0.3$ or $\tau^* < 0.2$, the cloud decision images were (by visual inspection) unsatisfactory. Therefore, we decided to limit τ^* in the interval $0.2 \leq \tau^* \leq 0.3$. The second modification includes making the adaptive thresholds localized depending on masks designed to partition the entire image into four regions, as illustrated in Fig. 5. The different colored perimeters in the image in this figure represent the different ROIs of each of the four partitions. The first ROI (blue) is represented with the circular perimeter of 60 pixel radius with center at the sun's position. The second ROI (cyan) is represented by an arc subtending 90° with its sector originating from the center of the image, but not including pixels in the first ROI. The third ROI (yellow) is represented by the circle of radius of 100 pixels, but not including the first and second ROI. The fourth ROI, includes any pixels not covered by the first, second, and third ROI. Each local τ^* values are computed in the union of the individual partitioned ROI and the ROI derived in Section 3.1. The resulting cloud indexed images are then combined into a single cloud decision image (e.g., Fig. 1c).

6. Grid cloud fractions

In order to forecast solar irradiance, we first construct a set of grid elements originating from the sun's position and oriented in column-wise fashion in the reverse direction of the cloud flow field as shown in Fig. 6. The direction is obtained from the representative velocity vectors described in Section 4. With this methodology the number of grid elements is limited by the size of the image (255×255). In this work, up to 6 grid elements are used. For each grid

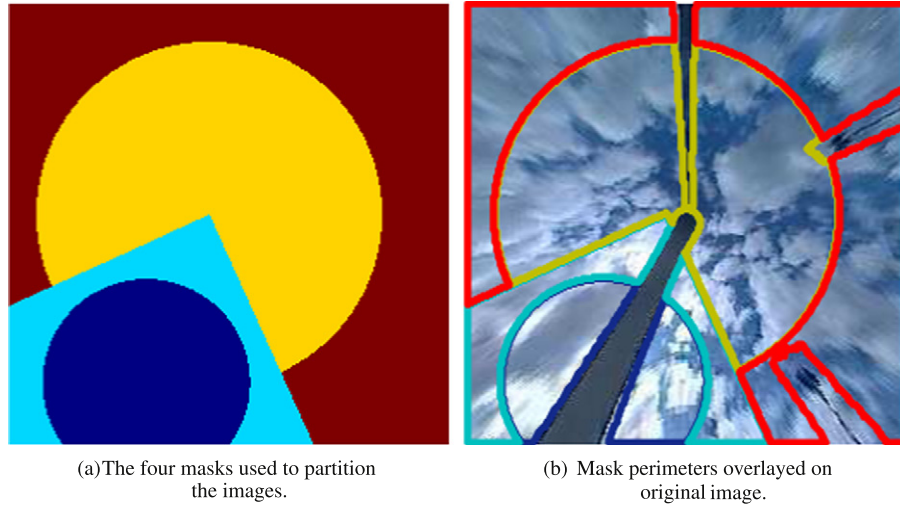


Fig. 5. Masks used for partitioning image into 4 regions for separate cloud classification thresholds. *Right*: Each color represents a different mask; M_1 : blue, M_2 : cyan, M_3 : yellow, M_4 : red. *Left*: Perimeters of masks overlaid on original image. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

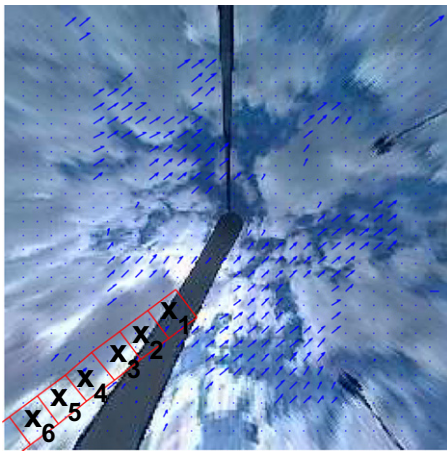


Fig. 6. The squares indicate regions of interest assigned for variables X_1, \dots, X_6 . Each variable represents the computed cloud fraction in the designated regions.

element, we compute the cloud fractions, denoted as X_i , in the area A_i enclosed by the grid's boundaries, e.g.,

$$X_i = 1/N_i \sum_{(x,y) \in A_i} B_\tau(x,y), \quad (11)$$

where N_i is the number of valid pixels and the area of each element is 20×20 pixels.

The orientation of the grid elements allows for making solar forecast predictions at several time-horizons. Grid elements further away from the sun's location on the image may be more useful for longer forecast horizons. For example, if the average cloud speed is 10 pixels/min, then the grid element which contains the best information for a 5 min forecasting horizon should be a distance of 10 pixels/min \times 5 min = 50 pixels. This distance corresponds to the third grid element whose midpoint is also 50 pixels from the sun position. Although the cloud speed is not accounted explicitly in the following applications, it is clear

which grid elements will likely be more useful for a certain forecast horizon.

The size of the grid elements may correspond to the time-averaging of the solar irradiance. Since we are computing 1-min averages, it can be argued that the size of the grid should be 10×10 pixels (assuming cloud speeds of 10 pixels/min). This size, however, was found to give forecasting results that were too erratic possibly due to errors in cloud pixel identifications. Therefore, we increased the grid element areas to 20 pixels \times 20 pixels, which gave more reliable results. Tuning the size of the grid areas to larger or smaller values did not yield any noticeable forecasting differences, so we deemed further optimization unnecessary here.

7. Forecasting direct normal irradiation

In the following calculations we selected four days to apply the DNI forecasting algorithm. The four days selected are given in Table 1. These days were selected for the nature of their cumuloform clouds (which facilitate cloud classification), but also for being highly variable DNI days. Cloud speeds and DNI attributes for the days under consideration are also summarized in Table 1.

7.1. Evaluation of errors

We test 1-min averaged direct normal irradiance forecasts for time horizons (Δt) of 3–15 min. To compare the cloud fractions with DNI values, each variable $X_i(t)$ is subtracted from 1 and then multiplied by DNI_{\max} ,

$$X_i(t)' = DNI_{\max}(1 - X_i(t)). \quad (12)$$

We subtract from 1 is because of the inverse relationship between cloud fraction and DNI intensity. DNI_{\max} is the maximum beam irradiance (DNI) value for the days

Table 1

Days used in forecasting evaluations with summary statistics including averages and standard deviations of cloud speeds, cloud directions, and DNI values.

Day	Ave speed (pixels/min)	Stdv speed (pixels/min)	Ave direction (°)	Stdv direction (°)	Ave DNI (W/m ²)	Stdv DNI (W/m ²)
June 1, 2011	6.06	0.86	56.25 (NE)	17.71	443	401
June 5, 2011	11.62	1.51	110.48 (NW)	11.66	571	384
October 5, 2011	8.06	2.06	−17.10 (SE)	13.43	572	343
November 4, 2011	7.84	4.48	−27.86 (SE)	38.35	631	310

analyzed, which is set to a constant value of 900 W/m² for this study. The errors are summarized by calculating Root-Means-Squared Errors (RMSEs) which are computed for each variable and for each forecasting horizon as

$$\text{RMSE}(\Delta t, i) = \sqrt{\frac{1}{N} \sum_{t \in T} (DNI(t + \Delta t) - X_i(t'))^2}, \quad (13)$$

where the summation is carried out in the set $T = 10:00\text{--}14:00$ PDT, corresponding to 4 h of the evaluated day. We also consider the persistence forecast for which the RMSE is

$$\text{RMSE}_p(\Delta t) = \sqrt{\frac{1}{N} \sum_{t \in T} (DNI(t + \Delta t) - DNI(t))^2}. \quad (14)$$

The persistence model is the simplest forecasting model available, and can be remarkably accurate (in the statistical sense) for short term horizons. Persistence represents a the

situation where no knowledge or estimation of future values is known, and only current deviations from clear sky conditions are considered. Also, as discussed in Marquez and Coimbra (2013), the RMSEs of the persistence model roughly represents the variability of the solar irradiance, as defined in Hoff and Perez (2010). Essentially, when the variability is low, an appropriately defined persistence model produces very accurate forecasts and therefore any forecast model which shows improvement over the persistence is improving the estimation over *random* variability (Marquez and Coimbra, 2013).

7.2. Forecast accuracy

Table 2 show the RMSEs as calculated by Eqs. (13) and (14) for each variable X_i and the persistence model. The forecasting horizons are denoted in the 1st column of the

Table 2

Root-Mean-Squared Errors (RMSEs) in W/m² computed for forecasting horizons of 3–15 min as indicated in 1st columns. Bold-faced numbers represent best RMSE with respect to time forecasting horizon. Generally, grid element variables located further away are more useful for longer forecast horizons. Maximum values of DNI for these days range from 800–900 W/m² as indicated in Fig. 8

Forecast Horizon	Persistence	X_1	X_2	X_3	X_4	X_5	X_6	Forecast Horizon	Persistence	X_1	X_2	X_3	X_4	X_5	X_6
(a) [June 1, 2011]								(b) [June 5, 2011]							
3	437	361	444	518	531	565	576	3	279	258	280	313	333	347	361
4	475	286	406	494	519	558	564	4	301	213	241	293	321	343	345
5	495	226	371	473	504	549	555	5	326	236	209	274	307	334	335
6	496	238	334	450	492	533	545	6	360	283	225	250	296	323	331
7	504	292	307	427	476	514	533	7	379	312	261	242	278	317	326
8	518	354	300	400	455	496	522	8	390	328	279	269	277	308	325
9	535	400	316	367	433	480	507	9	403	346	316	294	305	312	330
10	547	431	343	341	404	458	496	10	415	368	338	317	325	327	341
11	545	446	367	336	376	429	477	11	424	392	355	337	337	332	349
12	554	457	388	355	351	403	451	12	436	410	377	355	350	345	353
13	570	466	399	375	346	372	422	13	455	417	398	374	370	366	360
14	586	482	411	392	358	358	398	14	463	421	413	394	387	385	373
15	596	493	433	405	370	360	378	15	467	433	420	412	401	402	392
(c) [October 5, 2011]								(d) [November 4, 2011]							
3	352	276	264	302	331	375	377	3	309	237	329	395	416	408	408
4	376	317	252	310	319	360	374	4	322	217	312	392	415	414	414
5	401	360	267	306	318	348	366	5	338	257	302	383	412	420	420
6	412	399	293	294	318	341	359	6	352	286	304	375	411	428	428
7	427	421	330	289	314	348	358	7	361	292	305	366	409	440	440
8	442	445	361	305	307	356	367	8	368	298	299	359	396	437	437
9	451	468	389	322	300	355	370	9	375	305	291	360	386	433	433
10	453	479	409	345	303	344	372	10	390	311	283	356	380	430	430
11	453	487	422	359	317	341	367	11	412	315	275	339	373	424	424
12	451	496	424	361	334	332	362	12	433	320	267	317	371	42	425
13	454	497	431	353	347	328	349	13	447	336	276	297	365	426	426
14	463	496	433	352	352	331	340	14	461	358	291	291	359	424	424
15	475	496	438	356	356	335	332	15	473	380	307	299	354	424	424

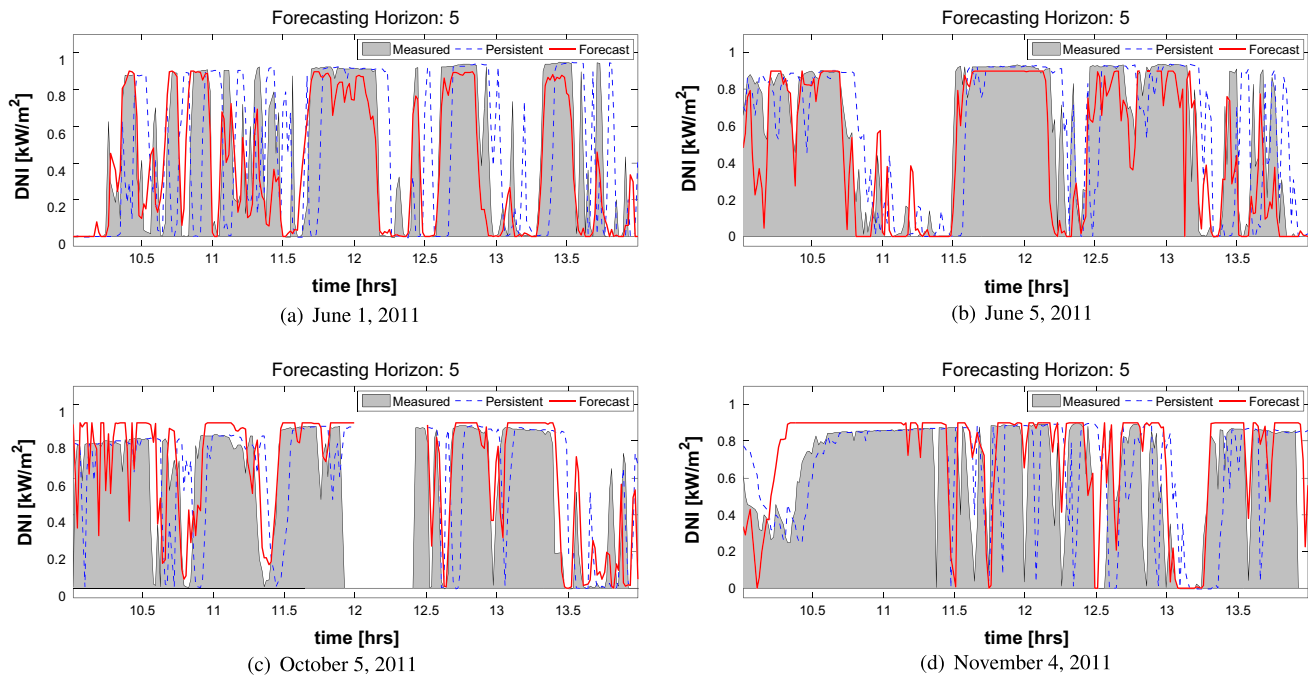


Fig. 7. 5-Min ahead forecasting of 1-min averaged DNI values for the 4 days evaluated. The time-series range from 10:00 to 14:00 PDT. Value between 12:00 to 12:30 were removed because of poor cloud classification of cirrus clouds.

table. Each sub-table corresponds to a different day. According to each of the sub-tables, there is a clear trend in the variable performance and the time horizon. Variables representing grid elements further away from the solar region are more useful for predicting DNI at longer time horizons. The most accurate forecasts appear to be for 5-min ahead, probably because of the relatively short forecast horizon and because the grid elements that correspond to the best 5-min ahead forecasts are minimally affected by the glaring of the sun.

The DNI forecasting errors are evaluated further by making comparisons with the errors of the persistence model. Time-series plots of the measured values, the persistence model, and the 5-min ahead forecasts with lowest overall error are shown in Fig. 7. We can see from these figures that the forecasts show considerable improvement over the persistence to predict the onset of the effect of cloud transient on DNI. During periods where there is high frequency of fluctuations, the forecasts do not overlap the measured values as well and are therefore still difficult to predict.

By taking the RMSE data of the best models and the persistence model from Table 2, we produced Fig. 8. This figure, again shows that the optimal forecast period is for 5 min ahead. For each day analyzed, there are significant improvements over the forecast period with RMSEs, on average, at approximately 300 W/m^2 . To get a better indication of the quality of the forecast models, we display in Fig. 9 the same RMSEs in Fig. 8, but this time plotting the RMSEs of the forecast versus the RMSEs of the persistence model. Although there is some scatter, a general trend is apparent, which indicates that the forecast RMSEs

increases with the persistence model RMSEs. In analogy with Marquez and Coimbra (2013), the ratio of the RMSEs is approximated with the slope of a linear fit to the scatter data. In this case, we approximate the slopes to range between 0.6 and 0.8, which means that the forecast quality of the models as a measure of improvement over a persistence model are between $1 - \text{RMSE}/\text{RMSE}_p \approx 1 - 0.8 = 20\%$ and $1 - \text{RMSE}/\text{RMSE}_p \approx 1 - 0.6 = 40\%$, where higher values indicate better forecasting quality. The improvement in forecasting skill is competitive in terms of other forecasting models described in the literature for different time horizons. For 1-h ahead forecasting, Perez et al. (2010) obtained 8–10% improvement over

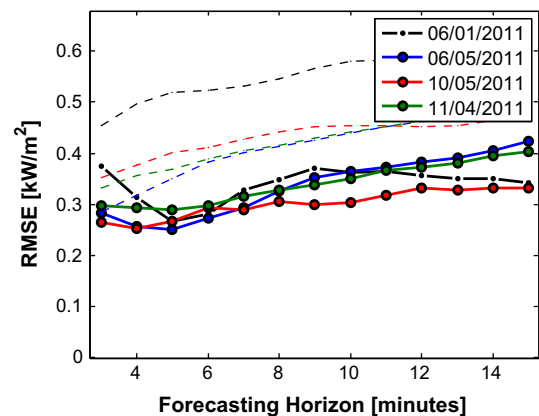


Fig. 8. RMSEs and forecast horizons. Bold lines with markers denote the best forecast RMSE for the forecast horizon and dashed thin lines denote the persistent RMSEs. The optimal forecast horizon is found consistently around the 5-min ahead mark.

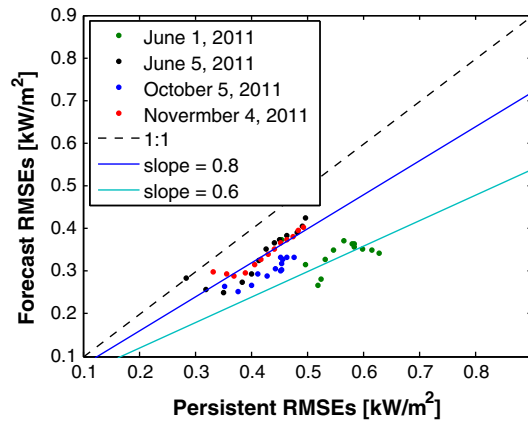


Fig. 9. Comparing *best* RMSEs with persistence model RMSEs to evaluate ‘forecast skill’. The slope of 0.8 suggests that the forecast skill is between $1 - RMSE/RMSE_{persistence} \approx 1 - 0.8 = 20\%$ and $1 - RMSE/RMSE_{persistence} \approx 1 - 0.6 = 40\%$. See Marquez and Coimbra (2013) for a detailed discussion on this metric for evaluating solar forecasting methods.

persistence using cloud motion forecasts from satellite images; Marquez and Coimbra (2013) obtained 12–20% using a ANNs; and Martin et al. (2010) obtained 10–15% also using ANNs. Unfortunately, direct comparisons are difficult to make due to regional meteorological conditions and different time horizons. Nonetheless, our preliminary study shows that within the 15 min time horizon, forecasting with sky image processing shows promise, especially if coupled with stochastic learning techniques that can offset some of the image classification and translation issues described above.

8. Conclusions

The work presented here makes significant progress in our ability to forecast DNI for time horizons shorter than 15 min. Procedures for image processing were applied, including the development of a spatial transformation to map the image obtained by a hemispherical mirror to locations on a flat grid, application of a PIV routine (Mori and Chang, 2003) to compute velocity fields and the determination of a representative velocity vector using a k -means algorithm, and a modified application a technique developed by Li et al. (2011) for cloud pixel classification. We introduced an approach for computing cloud fractions for grid elements oriented in such a way to scan the image according the computed velocity vectors. The computed cloud fractions were used for forecasting DNI. Evaluations of the RMSE errors demonstrate that sky imagers are useful for 3–15 min ahead forecasting horizons. Comparing with a persistent model, it appears that the most significant forecast accuracies are for 5-min ahead.

We chose four highly variable days to demonstrate the methodology. While four days may seem too small a data set for validation, a limited number of days is in fact more relevant to evaluate short-term forecasts than a very large period including many clear day periods because clear days

tend to affect the error averages (there is essentially no forecast errors for either clear days or consistently cloudy days). We have also performed longer-term (up to 6 months) evaluations of this methodology that confirm the error statistics presented in greater detail in this work.

Results of the present solar forecasting method are very encouraging despite some difficulty in cloud identification and the inability to correctly account for multiple layers of clouds moving in distinct directions. Previous work using sky imagers highlighted intrinsic difficulties with achieving robust image cloud classifications (Long et al., 2006; Chow et al., 2011; Li et al., 2011; Crispim et al., 2008; Huo and Lu, 2009), in particular for images with large amounts of glazing. We experienced similar difficulties with this work. We anticipate that improved cloud classification of images plus the incorporation of stochastic learning techniques for translation error reduction will contribute significantly to forecasting accuracy at short term horizons.

The methodology described here may not be strictly applicable to cases where distinct cloud layers move in different directions, especially if these situations occur frequently. Nonetheless, remote sensing-based forecasts also suffer from the same type of multi-layer limitation. The sector method we propose capitalizes on the ability to identify the preferential cloud motion and process the information contained in the images expeditiously, and has the potential to create short-term forecasts that perform substantially better than persistence forecasts. The assumption of single steady state cloud cover is considered one of the main weaknesses of image-based solar forecasts (Krostylev and Povlovski, 2012), but this can be potentially circumvented by stereo-photography systems with increased resolution and dynamic range.

Acknowledgements

The PI of this project (CFMC) gratefully acknowledges the partial financial support given by the California Energy Commission (CEC) under the PIER RESCO Project PIR-07-036, which is managed by Mr. Hassan Mohammed; and by the National Science Foundation (NSF) EECS (EPAS) award No. 1201986, managed by Dr. Paul Werbos. We also acknowledge the Eugene Cotta-Robles (ECR) Fellowship program of the University of California and Southern California Edison Fellowship program funding for RM. Seed and continued support by the Center for Information Technology Research in the Interest of Society (CITRIS) is also gratefully appreciated. Special acknowledgments to Mr. David Larson of the Solar Forecast Engine Laboratory at UCM/UCSD for maintaining instrumentation and databases used in this work.

References

- Cao, J., Lin, X., 2008. Study of hourly and daily solar irradiation forecast using diagonal recurrent wavelet neural networks. *Energy Conversion and Management* 49 (6), 1396–1406.

- Changsong, C., Shanxu, D., Tao, C., Bangyin, L., 2011. Online 24-h solar power forecasting based on weather type classification using artificial neural network. *Solar Energy* 85 (11), 2856–2870.
- Chow, W.C., Urquhart, B., Lave, M., Dominguez, A., Kleissl, J., Shields, J., Washom, B., 2011. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. *Solar Energy* 85 (2011), 2881–2893.
- Crispim, E.M., Ferreira, P.M., Ruano, A.E., 2008. Prediction of the solar radiation evolution using computational intelligence techniques and cloudiness indices. *International Journal of Innovative Computing, Information and Control* 4 (5), 1121–1133.
- Duffie, J.A., Beckman, W.A., 2006. *Solar Engineering of Thermal Processes*, third ed. John Wiley & Sons Inc., Hoboken, New Jersey.
- Hammer, A., Heinemann, D., Lorenz, E., Ckehe, B.L., 1999. Short-term forecasting of solar radiation: a statistical approach using satellite data. *Solar Energy* 67 (October), 139–150.
- Hoff, T.E., Perez, R., 2010. Quantifying pv power output variability. *Solar Energy* 84 (10), 1782–1793.
- Huo, J., Lu, D., 2009. Cloud determination of all-sky images under low-visibility conditions. *Journal of Atmospheric and Oceanic Technology* 26 (10), 2172–2181.
- Krostylev, V., Povlovski, A., 2012. Solar power forecasting performance – towards industry standards. In: 92nd American Meteorological Society Annual Meeting <<http://ams.confex.com/ams/92Annual/webprogram/Paper203131.html>>.
- Li, Q., Lu, W., Yang, J., 2011. A hybrid thresholding algorithm for cloud detection on ground-based color images. *Journal of Atmospheric and Oceanic Technology* 28 (Oct), 1286–1296.
- Long, C.N., Ackerman, T.P., Gaustad, K.L., Cole, J.N.S., 2006. Estimation of fractional sky cover from broadband shortwave radiometer measurements. *Journal of Geophysical Research* 111 (D11204).
- Lorenz, E., Hurka, J., Heinemann, D., Beyer, H.G., 2009. Irradiance forecasting for the power prediction of grid-connected photovoltaic systems. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 2 (1), 2–10.
- Marquez, R., Coimbra, C.F.M., 2011. Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database. *Solar Energy* 85 (5), 746–756.
- Marquez, R., Coimbra, C.F.M., 2013. Proposed metric for evaluation of solar forecasting models. *Journal of Solar Energy Engineering* 135 (011016), 9.
- Marquez, R., Gueorguiev, V.G., Coimbra, C.F.M., 2013. Forecasting of global horizontal irradiance using sky cover indices. *Journal of Solar Energy Engineering* 135 (011017), 5.
- Martin, L., Zarzalejo, L.F., Polo, J., Navarro, A., Marchante, R., Cony, M., 2010. Prediction of global solar irradiance based on time series analysis: application to solar thermal power plants energy production planning. *Solar Energy* 84 (10), 1772–1781.
- Mathiesen, P., Kleissl, J., 2011. Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States. *Solar Energy* 85 (5), 967–977.
- Mellit, A., 2008. Artificial intelligence techniques for modelling and forecasting of solar radiation data: A review. *International Journal of Artificial Intelligence and Soft Computing* 1, 52–76.
- Mellit, A., Pavan, A.M., 2010. A 24-h forecast of solar irradiance using artificial neural network: application for performance prediction of a grid-connected pv plant at Trieste, Italy. *Solar Energy* 84 (5), 807–821.
- Mellit, A., Eleuch, H., Benghanem, M., Elaoun, C., Pavan, A.M., 2010. An adaptive model for predicting of global, direct and diffuse hourly solar irradiance. *Energy Conversion and Management* 51 (4), 771–782.
- Mori, N., Chang, K.-A., 2003. Introduction to MPIV <<http://www.ocean-wave.jp/software/mpiv>>.
- Neto, M., Luiz, S., von Wangenheim, A., Pereira, E.B., Comunello, E., 2010. The use of euclidean geometric distance on rgb color space for the classification of sky and cloud patterns. *Journal of Atmospheric and Oceanic Technology* 27 (9), 1504–1517.
- Peng, T., 2011. Detect circles with various radii in grayscale image via Hough Transform. *Matlab File Exchange* <<http://www.math-works.com/matlabcentral/fileexchange/9168>>.
- Peng, T., Balijepalli, A., Gupta, S.K., LeBrun, T., 2007. Algorithms for on-line monitoring of micro spheres in an optical tweezers-based assembly cell. *ASME Journal of Computing and Information Science in Engineering* 7(4), 330–338.
- Perez, R., Kivalov, S., J., S., Hemker, K.J., Renne, D., Hoff, T.E., 2010. Validation of short and medium term operational solar radiation forecasts. *Solar Energy* 84, 2161–2172.
- Reikard, G., 2009. Predicting solar radiation at high resolutions: a comparison of time series forecasts. *Solar Energy* 83 (3), 342–349.
- Sfetsos, A., Coonick, A.H., 2000. Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques. *Solar Energy* 68 (2), 169–178.
- YES, 2011. TSI-880 Automatic Total Sky Imager Manual. Yankee Environmental Systems, Inc., Airport Industrial Park 101 Industrial Blvd. Turners Falls, MA 01376 USA.