Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs

Ricardo Marquez [a], Hugo T.C. Pedro [b], Carlos F.M. Coimbra [b,*]

[a] *Mechanical Engineering and Applied Mechanics Program, School of Engineering, University of California Merced, Merced, CA 95343, USA*
[b] *Department of Mechanical and Aerospace Engineering, Jacobs School of Engineering, and Center of Excellence in Renewable Energy Integration and Center of Energy Research, University of California San Diego La Jolla, CA 92093, USA*

Abstract

This work describes a new hybrid method that combines information from processed satellite images with Artificial Neural Networks (ANNs) for predicting global horizontal irradiance (GHI) at temporal horizons of 30, 60, 90, and 120 min. The forecast model is applied to GHI data gathered from two distinct locations (Davis and Merced) that represent well the geographical distribution of solar irradiance in the San Joaquin Valley. The forecasting approach uses information gathered from satellite image analysis including velocimetry and cloud indexing as inputs to the ANN models. To the knowledge of the authors, this is the first attempt to hybridize stochastic learning and image processing approaches for solar irradiance forecasting. We compare the hybrid approaches using standard error metrics to quantify the forecasting skill for the several time horizons considered.
© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Accommodating higher penetration levels of solar power into a new generation of power grid portfolios requires the use of increasingly more accurate forecasting systems in order to reduce backup reserves and improve unit commitment associated with power generation variability (Lew and Piwko, 2010; California Independent System Operator (CAISO), 2010). The objective of this work is to develop solar irradiance forecasts for time horizons in the range of 30 min to 2 h. We develop solar forecasts for two locations in the Central Valley of California using information extracted from GOES-West satellite images, which are then processed to be optimized by Artificial Neural Networks (ANNs).

The most important factor affecting the amount of global solar irradiation at any given time of a day in California's central valley is cloud cover. Predicting accurately the extent, motion, formation, dissipation, and transmittance of ever deforming clouds is a complex and somewhat unrealistic task for solar forecast applications, even for moderate temporal (few minutes ahead) and spatial scales (few kilometers). As a result, the efficacy of short-term satellite-based forecasting of solar irradiance is penalized when the models are not robust to the uncertainty in the many parameters involved. For this reason, we developed a method to combine ANNs with detailed satellite image processing for determining future values of global horizontal irradiance (GHI).

Solar forecasting is a timely topic, and several short-term solar irradiance forecasting approaches have been presented recently (Marquez et al., 2013; Martin et al., 2010; Marquez and Coimbra, 2011; Cao et al., 2008;

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

Nomenclature

α, a_n	empirical fitting parameters
$\overline{\mathbf{V}}, \overline{\mathbf{V}}'$	average velocity and smoothed average velocity
Δt	time-series time step
$\widehat{G}_{f,t+k\|t}$	forecasted GHI (kW/m^2) using model f for forecasting horizon $k\Delta t$
$\mathbf{u}$	exogenous variables
ρ	albedo
θ, θ_t	sun's zenith angle, sun's zenith angle at time t (rad)
E	pixel intensity for images of the visible channel
G	measured GHI (kW/m^2)
G_{clr}	GHI clear-sky model
IR	pixel intensity for images of the Infra-Red channel
k^*	clear-sky index
n	cloud index
U, V	uncertainty and variability of the time-series
u, v	horizontal (longitudinal) velocity and vertical (latitudinal) velocity
X_i	cloud fraction for grid element i
ANN	Artificial Neural Networks
CMF	cloud motion forecast
GHI	global horizontal irradiance (kW/m^2)
GOES	Geostationary Operational Environmental Satellites
MBE	mean bias error
MSE	mean squared error
PIV	particle image velocimetry
RMSE	root mean squared error

Bacher et al., 2009; Mellit et al., 2010; Reikard, 2009). However, direct use of satellite-imagery for solar forecasting is less widespread due to the available temporal and spatial resolutions. These models typically rely on cloud translation and use the clear-sky index to forecast solar irradiance (Hammer et al., 1999b, 2001, 2003), (Hammer et al., 1999a, 2001, 2003). Many of the methodologies that are not satellite-based rely heavily on stochastic learning techniques such as ANNs as the main forecast method (Marquez et al., 2013; Marquez and Coimbra, 2011; Martin et al., 2010; Cao et al., 2008; Bacher et al., 2009; Mellit et al., 2010), with another set of authors using auto-regressive models (Bacher et al., 2009; Martin et al., 2010; Reikard, 2009). The satellite-based work of Hammer et al. (2001) uses ANNs but without incorporating into the stochastic learning model the results of the image-processing used for the cloud translations. The aim of this study is to combine the ANNs with the detailed information gathered through the image processing stages.

This work involves the development of two forecasting models. The first GHI forecast model only includes satellite-derived information such as the cloud fractions computed at prescribed grid elements of the satellite images. This model complements the recent study of Pedro and Coimbra (2012) where only endogenous inputs were considered in order to examine the effectiveness of regression and stochastic learning approaches. The second GHI forecast model is a generalization over the first model since lagged values of GHI are used in addition to the satellite-derived inputs to examine the effectiveness of remote sensing information in improving forecast skill.

As the energy industry moves towards the integration of increasing levels of variable renewable resources, several authors (e.g., (Marquez and Coimbra, 2013b; Kostylev and Pavlovski, 2011)) point out the need for improved comparison metrics and best practices for comparison of different forecasting methods. Therefore, a secondary goal

of this work is to compare in detail the results of the models described above with other satellite-based solar forecasting methods such as in Hammer et al. (2003), Perez et al. (2010b).

The organization of this paper is as follows: first we present an overview of the methodology in Section 2, which discusses the modeling methodology as composed of three distinct stages. In Sections 3 and 4, the satellite-image processing stage that includes cloud indexing and image velocimetry is explained. Section 5 describes the methodology of building and incorporating input variables for the ANN forecasting models. Details regarding irradiance data are explained in Section 6. The configuration and training of the ANN models are described in Section 7. The persistence model used as the baseline for the evaluation of the forecasting performance is presented in Section 8. Finally, the different models are evaluated in Section 9 and the main concluding remarks are provided in Section 10.

2. Forecast model overview

The methodology we propose in this work capitalizes on combining the strengths of ANNs for time-series predictions with the relevance of information available in remote sensing images. An ANN forecasting method that uses information of lagged values of single or multivariate time-series can be written generically as

$$\widehat{G}_{t+k|t} = f(G_t, G_{t-\Delta t}, \cdots, \mathbf{u}_t, \mathbf{u}_{t-\Delta t}, \cdots), \tag{1}$$

where G is the endogenous variable, and $\mathbf{u}$ is a vector of exogenous variables. The lagged values are considered inputs into the forecasting model which has an output $\widehat{G}_{t+k|t}$, where $k|t$ denotes the forecasting horizon given in multiples of the time-series time step Δt. In this work we consider fixed forecasting horizons of 30, 60, 90, and 120 min for a time-series G_t consisting of 30-min averaged GHI ($k = 1, 2, 3$ and 4 respectively). Therefore we cover

time compression ratios (time horizon over averaging period) ranging from 1 to 4.

A critical step in developing the forecasting model is a careful consideration of which data to include in the input vectors. Ideally, the inputs should be those which contain information relevant to the future values of the time-series. In the case of solar irradiance, its future values are strongly dependent on cloud conditions, therefore a good characterization of cloud conditions and associated cloud motions is of critical importance in predicting ground levels of solar irradiance. Thus, in this work we propose a forecasting methodology for solar irradiance based on satellite cloud images that involves a three-stage approach. A schematic depicting an overview of the procedures involved in the model development is shown in Fig. 1. The first two stages involve procedures related to cloud analysis, including cloud indexing and cloud velocimetry. These two procedures can and should be done in parallel. In the stage identified by the upper-left boxes in Fig. 1, cloud indexed images are produced using images derived from the visible part of the solar spectrum. The stage identified by the upper-right boxes shows the velocimetry calculations using images obtained from the infrared (IR) channels. The output of this stage is a single velocity vector that represents the mean direction of cloud propagation. In the last stage, identified by the lower boxes, the cloud indexed image and the velocity vector are used to define a set of grid elements, and to compute the associated cloud fractions. The cloud

fraction variables define the inputs that will be used by the ANN forecasting method, which in turn produces the final solar irradiance forecast. Each stage in the forecast method is discussed in detail in the following sections.

3. Cloud indexed images

Cloud indexed images are produced by analyzing each pixel in the satellite images. The observed pixel intensity values of a raw visible image, E, measure the reflected sun light on the remote satellite sensors. Since clouds have a high reflectivity they appear whiter (due to the higher pixel intensity values) than the land masses and oceans in a satellite image. Snow on the ground also has high albedo and can therefore be easily confused with clouds when using a pixel based algorithm for cloud detection.

The cloud indexed image is obtained from E images following the approach described by Cano et al. (1986). This approach is described below because of the importance of this process in our methodology. The first step in the cloud indexing is to examine the pixel values as a function of the solar zenith angle $\theta(t)$. As described in Cano et al. (1986), we apply the following relation between pixel intensity values $E(t)$ and $\cos(\theta(t))$,

$$E(t) = I_0 \rho \cos(\theta(t))^{(1+\alpha)}, \tag{2}$$

where I_0 is the solar constant, ρ is the pixel albedo and α is an empirical parameter that characterizes the total atmo-
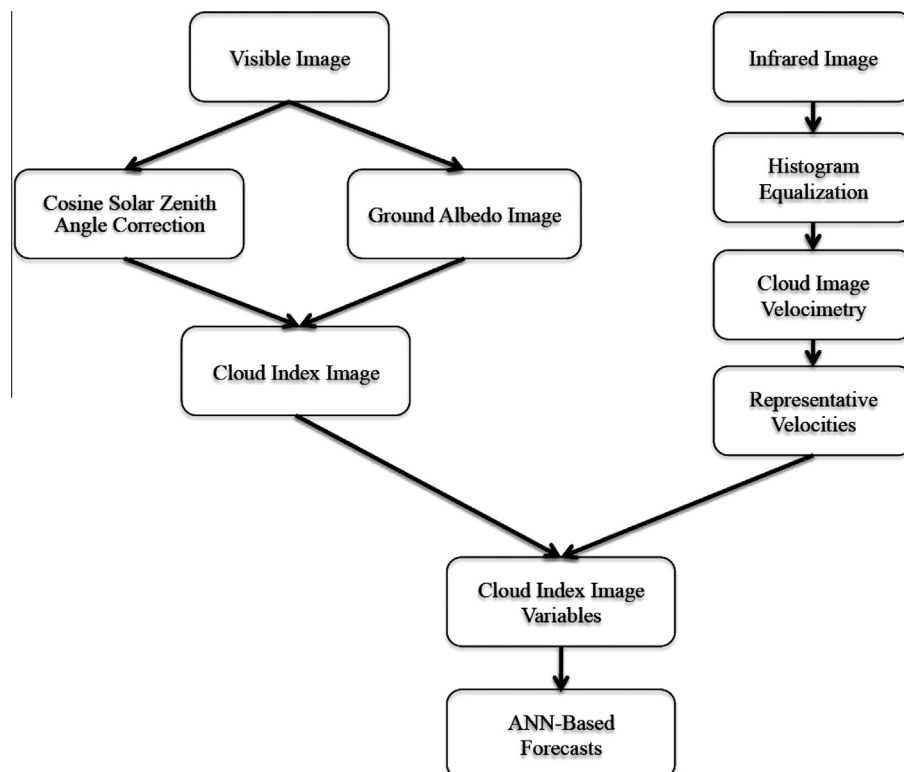

Fig. 1. Schematic of solar forecast modeling methodology highlighting 3 main stages–cloud indexing, image velocimetry, and ANN predictions.
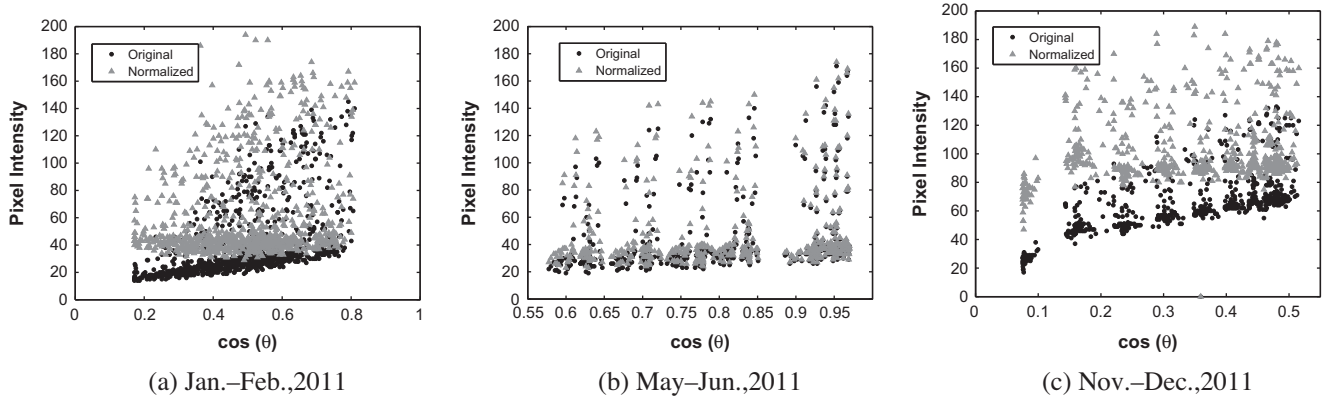
Fig. 2. Pixel intensity versus cos $(\theta(t))$ before and after normalization. The flatter line resulting from the normalization indicates that the diurnal effects on intensity have been removed.

spheric transmittance and depends upon the geographical location. Fig. 2 displays observed pixel intensity values versus $\cos(\theta(t))$ for a random location over several days indicated in the figure caption. The figure shows that the values of pixel intensity of the original image increase with $\cos(\theta(t))$ from values between 20 and 200. The diurnal variation of the solar irradiance is caused by $\cos(\theta(t))$ dependence of the solar irradiance, in order to remove the diurnal effects we use the following form of Eq. (2),

$$\rho = \frac{E(t)}{I_0 \, \cos(\theta(t))^{(1+\alpha)}}, \tag{3}$$

with parameter $\alpha = 0.4$ (determined by trial-and-error). In addition to the $E(t)$ pixel values (black markers), Fig. 2 shows values of $\rho \times I_0$ as a function of $\cos(\theta(t))$. The plots shown that the diurnal effects are successfully removed using $\alpha = 0.4$ for $\cos(\theta(t)) > 0.15$. However, for cases where $\cos(\theta(t)) < 0.15$ (close to sunrise and sunset) the diurnal effects are not removed in a satisfactory manner. As a result, in later sections, we only consider forecast analysis involving $\cos(\theta(t))$ values above 0.15.

After determining the normalized albedo image ρ, we can define a cloud index n as

$$n = \frac{\rho - \rho_{\max}}{\rho_{\max} - \rho_{\min}}, \tag{4}$$

where $\rho_{\max}$ and $\rho_{\min}$ are the maximum and minimum albedo determined over a set of images. In this work, $\rho_{\min}$ is obtained by taking the minimum ρ for each pixel location over a week's worth of satellite image data. These values are then assembled in a reference ground albedo image as exemplified in Fig. 3. The parameter $\rho_{\max}$ is not treated the same way as it depends on cloud reflectivity, depth and type which cannot be characterized properly by a reference image. In this work, the value of $\rho_{\max}$ is set to $\rho_{\max} = 200$. This value was selected based on the data observed in Fig. 2, whose pixel values do not exceed 200. In order to illustrate the cloud indexing procedure, Fig. 3 shows the reference ground albedo image, an unprocessed original image, and a cloud indexed image (using Eq. 4) for the central valley of California and part of the Sierra Nevada mountain range. The image pixel value limits in Fig. 3a and b are set to 80–200 for enhanced visualization. We can see from this figure how this procedure enhances the contrast between clouds and the ground. In the eastern part of the images around the Sierra Nevada, snow is observed and is not canceled out properly because of the
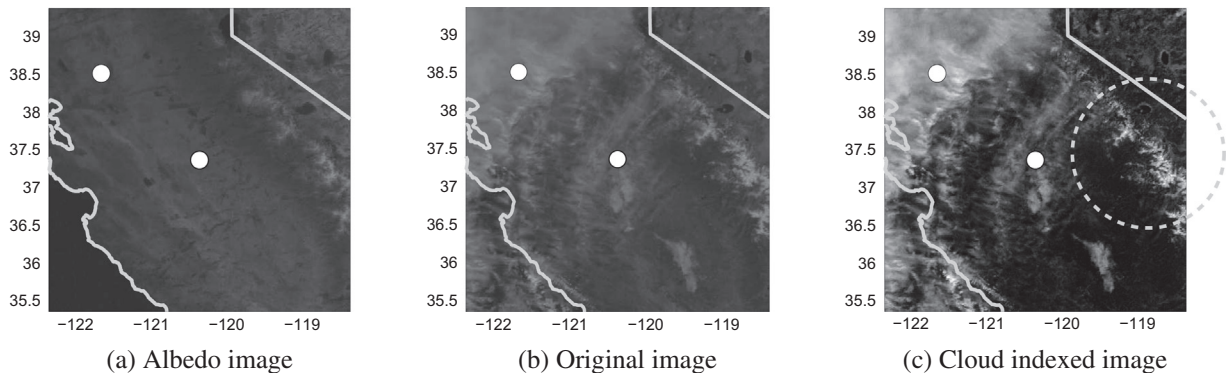

Fig. 3. An example of the cloud indexing procedure. *Left* Reference ground albedo image obtained by taking the minimum pixel intensities for each pixel location. *Middle* Original image taken on 11/23/2011. *Right* Cloud indexed image after applying normalization. The two dots identify the forecasting locations. *North-most* Davis. *South-most* Merced.

day-to-day snow accumulation during the time the images were taken, however this situation is not a concern for our purposes because there is no snow accumulation in the locations where the forecast will be applied.

The cloud indexed image in Fig. 3c is the output of the cloud processing stage (see upper-left boxes in Fig. 1). In a following stage we compute the cloud fractions on grid elements selected depending on the cloud motion. These variables will then be used in the ANN forecasting models.

4. Velocimetry with IR images

Velocity field calculations are performed using a Particle Image Velocimetry (PIV) algorithm obtained from Mori and Chang (2003). This algorithm has been successfully applied to the estimation of cloud motion from sky images obtained with a Total Sky Imager to produce Direct Normal Irradiance (DNI) forecasts (Marquez and Coimbra, 2013a). The PIV algorithm computes the velocity vectors by determining the correlation between a set of consecutive images separated by Δt. The images are partitioned in a grid of interrogation windows over which the algorithm operates. The PIV algorithm returns a velocity vector for each one of the interrogation windows.

A problem with using the images from the visible spectral bands (E) is that these images are only useful during the day. Moreover, they are difficult to process during the early and late parts of the day, as discussed in the previous section. To overcome this problem we proceed by using Infrared (IR) images since the IR images are available at all times and do not have the processing complications that would affect the PIV calculations if the visible images were used.

Given that the PIV algorithm applies correlations to neighboring pixels of sequences of images, if the pixel values change by factors other than the cloud motion the PIV algorithm will fail to capture the cloud apparent velocity. One possible source of such errors is the diurnal variation of the pixel intensity (see e.g., (Schillings et al., 2004)). Thus, in order to reduce potential errors using the PIV algorithm, each pair of IR images is first pre-processed by adjusting the intensity values of the second image (IR_t) with the first one ($IR_{t-\Delta t}$) using a histogram equalization algorithm. A histogram of an unprocessed image is shown in Fig. 4 (the black vertical bars). The distribution of pixel intensity for this example falls in the range 100–200 and is most dense at 120. Due to the narrow range, the original image appears mostly gray with little contrast between clouds and the ground in Fig. 5a. The image contrast is enhanced by adjusting the pixel value histograms so that the pixel values spread out between 0 and 255. An example of the pre-processing is shown in Fig. 5b. In the processed IR image, clouds are easily distinguished from the ground. The image that was taken 30-min later is also processed so that the histogram of the pixel intensity values matches the processed pixel values of the current image, see Fig. 5c. The processed consecutive IR images are used as
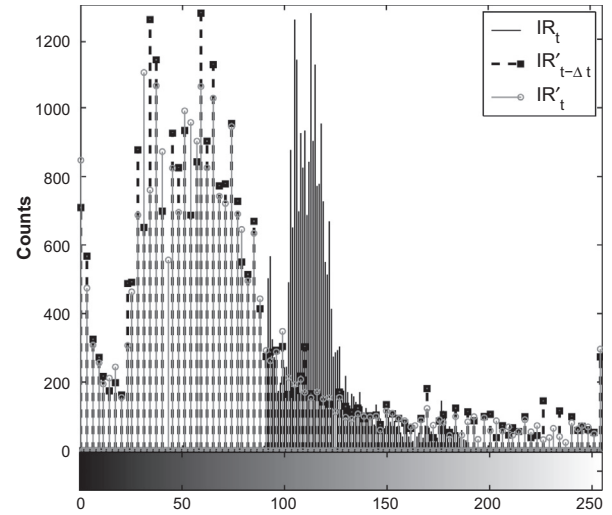

Fig. 4. Infrared image histograms adjusted so that errors due to the diurnal variation of the pixel intensity are not introduced into the PIV routine. By applying histogram equalization, the diurnality of the pixel intensities are removed.

inputs in the PIV algorithm to compute the velocity field vectors.

An example of the velocity field calculations is shown in the top plot of Fig. 6. The velocity field was obtained for images from the same days as the images in Fig. 5, from which we can see that the clouds are moving in the South-East direction. Although, there are a few 'stray' vectors in the PIV velocity field, we can see from the histogram of the velocity directions (rose plot on bottom-left of Fig. 6) and distribution of velocity vectors (compass plot on bottom-right of Fig. 6) that most vectors point in the South-East direction, as expected. To obtain the final output of the velocimetry stage (see Fig. 1), we compute the average velocity vector ($\overline{\mathbf{V}}$) as,

$$\overline{\mathbf{V}} = \frac{1}{N}\sum_{i\in\Im} u_i\mathbf{e}_{lon} + \frac{1}{N}\sum_{i\in\Im} v_i\mathbf{e}_{lat}, \tag{5}$$

where u is the speed in the horizontal (longitudinal) direction, v is the speed in the vertical (latitudinal) direction, and $\mathbf{e}_{lon}$ and $\mathbf{e}_{lat}$ denote the directional unit vectors. The summation is carried out over the set of vectors in $\Im \equiv \left\{i : \sqrt{u_i^2 + v_i^2} \geqslant 1\right\}$ (vectors with magnitude greater than 1 pixel per 30 min), so that only non-zero vectors are included in the averaging.

The final output of the velocimetry stage is $\overline{\mathbf{V}}_t$ which is computed from IR images at time t and $t - \Delta t$. Because at sub-hourly to hourly time-scales we do not expect that the bulk cloud movement will change very much, a backwards moving average over the last three time-steps is applied to ensure the smoothness of time-series $\overline{\mathbf{V}}_t$,

$$\overline{\mathbf{V}}'_t = \frac{1}{3}\sum_{i=0}^{2} \overline{\mathbf{V}}_{t-i\Delta t} \tag{6}$$

(a) Original IR at time t.　　(b) Processed IR at time $t - \Delta t$.　　(c) Processed IR at time t.
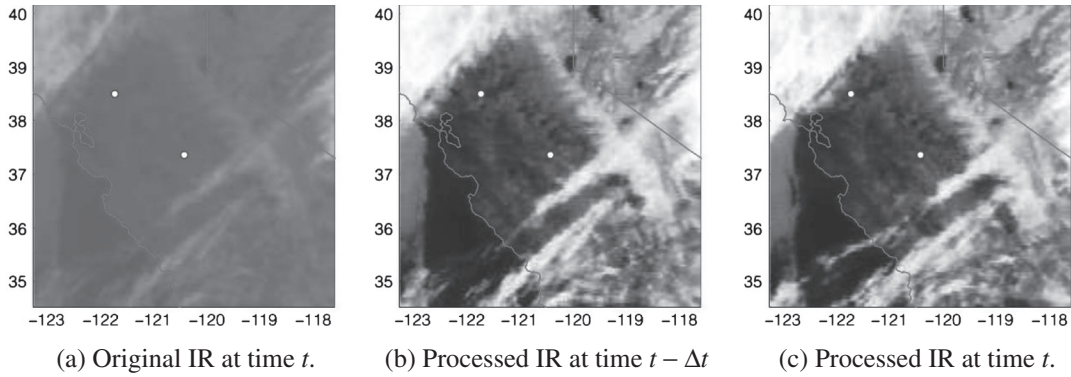
Fig. 5. Pre-processing of consecutive IR images so that histogram of pixel intensities are equalized. This step avoids potential errors in the PIV algorithm.
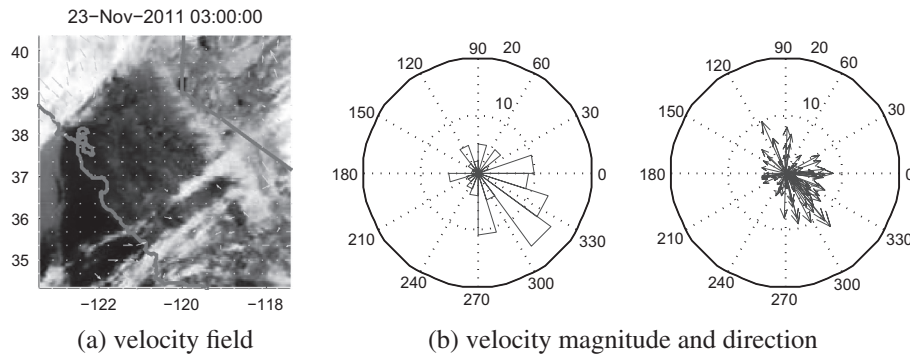

(a) velocity field　　(b) velocity magnitude and direction

Fig. 6. *Left:* velocity field obtained from PIV algorithm. *Middle:* rose plot shows the histogram of the velocity vectors; *Right:* the compass plot shows the distribution of the velocity vectors.

where $\overline{\mathbf{V}}_t'$ is the velocity that will actually be used for computing the set of cloud fraction variables that will be used as inputs for the ANN forecasting model.

For the majority of the time the cloud velocity computed from PIV represents well the velocity of all clouds in the image. However, in certain cases multiple layers of clouds moving in different directions at different speeds can occur. Since our methodology assumes a homogeneous cloud velocity and direction, these cases may result in large forecast errors. We did not remove from our analysis any periods where such circumstances occur. All the errors due to incorrect characterization of cloud speed and direction are included in the forecast performance data reported below.

5. Cloud fraction variables

As part of the first step in developing the forecast models, the input variables are established from the work done in the pre-processing stages. At the end of that stage there are two pieces of information available (1) cloud-indexed images, and (2) cloud velocities. The cloud fraction variables (X) are computed by creating a set of grid elements that represent portions of the sky that are more likely to determine the short term future of the solar irradiation at the forecasting location. These grid elements are created

with the first element originating at the location of interest, then extending out more elements in the upstream direction of the flow field as determined by $\overline{\mathbf{V}}_t'$. This approach assumes that advection is the main transport mode for the clouds, therefore it will likely perform well when frozen cloud translation is a good approximation.

Fig. 7 shows an example of a sector of elements for a sequence of images separated 2 time-steps apart (1-h). For each element the cloud fraction is computed using the cloud indexed image ($n(x,t)$ or $n_{x,t}$) obtained from the first stage (see Fig. 1). The calculation of X_i can be written as,

$$X_{i,t} = \frac{1}{N} \sum_{x_j \in A_i} n_{x_j,t}, \qquad (7)$$

where A_i represents pixels in the area bounded by the ith grid element (see Fig. 7). The size of the elements is adaptive and determined by analyzing the cloud speeds: higher speeds correspond to longer grid elements. A distribution of the cloud speeds is shown in Fig. 8 where the units are given in pixel widths (of the *IR* images) per 30 min. The average cloud speed is 3 pixels/30 min, which translates to $\approx$20–24 km/hour (the resolution of *IR* images is $\approx$4 km/pixel). Converting to units for the visible (E) images, which have a resolution of $\approx$1 km/pixel, the average speed is $\approx$12 pixels/30 min. The element widths are
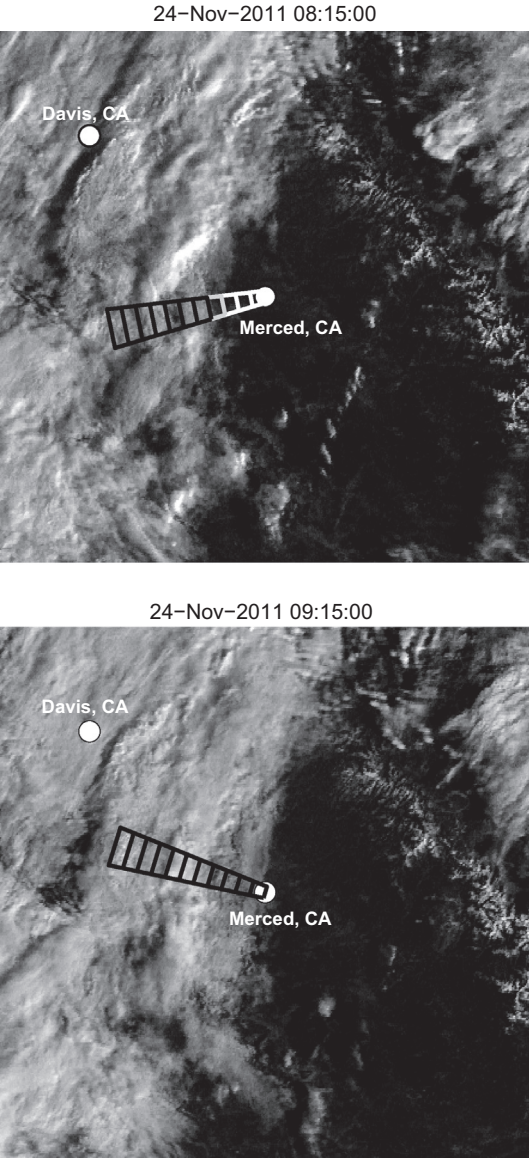
24−Nov−2011 08:15:00

24−Nov−2011 09:15:00

Fig. 7. Sector used for computing cloud fraction variables $X_{i,t}$. The sector orientation is collinear with the direction of the cloud movement. For this case, the sector originates from one of the locations of interest (Merced, CA). Another sector (not displayed in figure) for Davis, CA is also used.

adjusted to become wider for larger distances form the point of interest in order to have some robustness in all the uncertainty involved.

The cloud fraction variables just described and referred to as $X_1, X_2, \cdots, X_7$ are then used as exogenous inputs into the forecasting model. The subscripts denote the grid element for which the variable was computed with 1 being the grid element closest to the location of interest and 7 being the furthest. The approach described here resembles that in Marquez and Coimbra (2013a), except that, in that work the DNI forecasting was obtained with a simple curve fitting between the cloud fraction and the irradiance, whereas here these exogenous variables are incorporated into a non-linear ANN forecasting model.
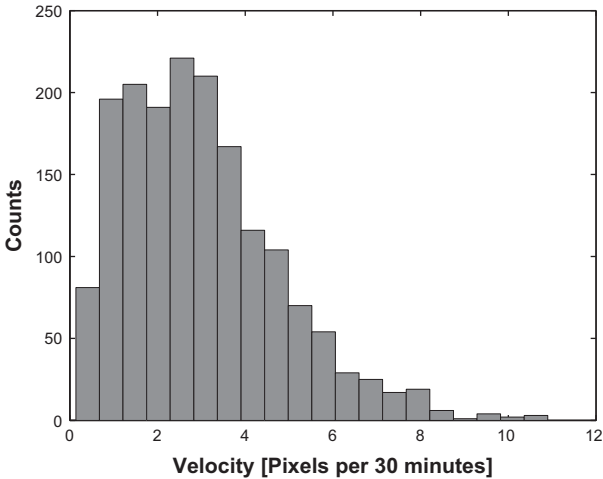

Fig. 8. Histogram of cloud speeds over the range of dates: Nov. 23, 2011–December 31, 2011. The velocity units are in pixels widths (of the IR images) per 30 min. The average cloud speed is 3 pixels/30 min, which translates to ≈ 20–24 km/h. Converting to units for the visible images, the average speed is ≈ 12 pixels/30 minutes.

6. Irradiance data

Together with satellite images this study uses GHI data for the two locations mentioned above gathered from January 1, 2011 through January 31, 2012. The raw data was captured at a 30 s. frequency and then averaged in 30 min. intervals. Data from the period between June 7, 2011 to November 22, 2011 was not included in the analysis because it consists almost exclusively of clear-sky days. One week worth of data from each month and each location was used as a hold-out testing set. The remaining data was used to train and validate the ANN models. Table 1 lists more details about the GHI data and Fig. 9 shows the distribution of GHI and the clear-sky index for the training/validation and the testing data sets. The histograms show that the variability conditions are very similar for both data sets.

7. Artificial neural networks

As mentioned in Section 2 and depicted in the forecasting algorithm diagram in Fig. 1, the forecasts are ultimately produced by an ANN time-series forecasting model. Artificial Neural Networks (ANNs) are useful tools for

Table 1
GHI data (30 min) used in this study and partition of the data between training/validation and testing sets.

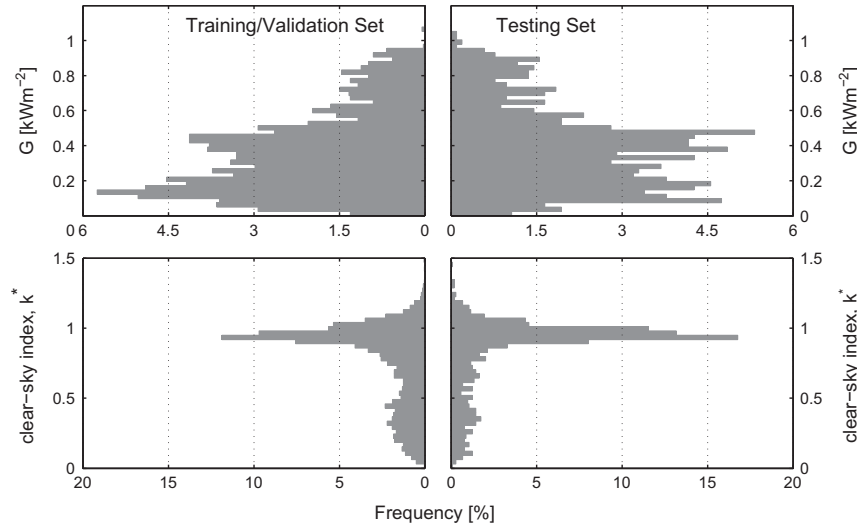| Location	Merced, CA	Davis, CA
Coordinates	37.36 N, 120.43 W	38.5 N, 121.73 W
Period of	01/01/2011 to 06/06/2011	11/23/2011 to
data collection	11/23/2011 to 01/31/2012	1/31/2012
Data Points	Total (both locations)	4788
($cos(\theta) > 0.15$)	Training/validation	3756 (78%)
	Testing	1032 (22%)

Fig. 9. GHI and clear-sky index k^* histograms for the training/validation and testing sets.

approximating complicated mapping functions for problems in classification and regression (Bishop, 1994), and have also been used extensively in many areas, including solar radiation modeling and forecasting (see e.g., (Pedro and Coimbra, 2012)). A review of over 40 applications of ANNs and other artificial intelligence techniques applied to solar radiation modeling can be found in Mellit (2008). The advantage of ANNs is that no assumptions are required about the underlying process relating input and output variables. An ANN is a particular representation of an output y in terms of some input variables x. The ANN representation is based on signals being sent through elements called neurons in such a way that the processing of the input signals produces an output, y, that is sufficiently close to the desired target value of t. Neurons are arranged in layers, where the first layer contains the set of inputs, the last layer contains the output, and the layers in between, referred to as hidden layers, contain hidden neurons. A feed-forward neural network with N inputs and N_h neurons in one hidden layer with a linear output activation function can be expressed as,

$$\widehat{G} = \sum_{i=1}^{N_h} w_i f_i^{hidden} \left(\sum_{j=1}^{N} w_{j,i} x_j + w_{0,i} \right) + w_0, \qquad (8)$$

where f_i^{hidden} are sigmoidal functions, such as the hyperbolic tangent function. A characteristic of feed-forward neural networks is that the neurons are successively interconnected from layer to layer where neurons in one layer affect all neurons in the next but do not affect other neurons in the same layer or any preceding layers. Numerical optimization algorithms such as back-propagation, conjugate gradients, quasi-Newton, and Levenberg–Marquardt have been developed to efficiently adjust the weights, $w_{j,i}$ and w_i in the feed-forward neural networks so that the minimization of some performance function is achieved on some training data. Typically the performance function used for adapting the weights is the mean squared error (MSE).

7.1. ANN forecast development and training

Here we use ANNs to produce the function to model the time-series function expressed in Eq. 1, which includes several types of input variables:

- irradiance: $G_t, \cdots, G_{t-n\Delta t}$,
- geometry: $\cos(\theta_{t+k\Delta t}), \cdots, \cos(\theta_{t-n\Delta t})$,
- cloud indices: $X_{i,t}, \cdots, X_{i,t-n\Delta t}$.

to predict $\widehat{G}_{t+k|t}$, where G are 30-min averaged values of GHI, $\cos(\theta)$ is also expressed in 30-min averages, and the time-series X_i's are as defined in Section 5. The models which are considered for this study are summarized in Table 2. The inputs include lagged information which is usually useful in forecasting models. The lagged information includes the previous $n = 2$ time-steps for the variables involved. A thorough analysis of the impact of lagged inputs was not performed, in part because of the computational effort necessary but mostly because the main objective of this work is to describe a new hybrid method that combines information from processed satellite images with artificial neural networks.

It is also important to include the time-series $\cos(\theta_t)$ as an input to the ANN, since this variable is useful for

Table 2
Model nomenclature and mathematical descriptions. Models $\widehat{G}_{f1}$ and $\widehat{G}_{f2}$ differ by their respective exclusion and inclusion of the G_t time-series as inputs. The Cloud Motion Forecast (CMF) model refers to the study in (Hammer et al., 2003), where $k^*_{f,t+k|t}$ is obtained by forecasting the cloud motion.

Model	Mathematical description
$\widehat{G}_{f1,t+k|t}$	$f(\cos(\theta_\tau)\|_{\tau=t+k\Delta t,\cdots,t-n\Delta t}, X_{i,\tau}\|_{\tau=t,\cdots,t-n\Delta t})$
$\widehat{G}_{f2,t+k|t}$	$f(\cos(\theta_\tau)\|_{\tau=t+k\Delta t,\cdots,t-n\Delta t}, X_{i,\tau}\|_{\tau=t,\cdots,t-n\Delta t},$
	$G_\tau\|_{\tau=t,\cdots,t-n\Delta t})$
$\widehat{G}_{p,t+k|t}$	$k_t^* G_{clr,t+k\Delta t}$
$\mathrm{CMF}_{t+k|t}$	$k^*_{f,t+k\Delta t} G_{clr,t+\Delta t}$

describing the deterministic diurnal variation of GHI. The variables X, if useful, will help to describe the cloud effects on GHI.

For this work we consider two ANN models with the main difference being in the input layer. The model that does not include lagged values of GHI as an input variable, is referred to as $\widehat{G}_{f1}$. The other model considers lagged GHI values as inputs and is referred to as $\widehat{G}_{f2}$. Otherwise, the ANN architectures for both models are configured to have $N_h = 5$ hidden neurons and are trained with quasi-Newton backpropagation that minimizes the MSE over a random training data set consisting of 85% of the data from the training/validation set (Table 1). The remaining 15% is used as a validation set in order to avoid ANN over-training by terminating the training once the MSE over the validation set has increased after 6 consecutive iterations. In order to account for the random selection of the training data and the random initialization of the quasi-Newton backpropagation algorithm each ANN-based model is trained 5 times and the associated results in the sections below are computed from the average of these runs.

8. Persistence model as baseline forecast comparison

In this section we explain the persistence model that is used as the baseline for comparing the accuracy of the forecasting methodology proposed here. The persistence model is defined by using the current clear-sky index as a prediction for the future clear-sky index. The clear-sky index is defined by normalizing GHI with a clear-sky model, G_{clr},

$$k^* = \frac{G}{G_{clr}}. \tag{9}$$

Assuming that the current k^* persists into the future, the forecasted GHI is:

$$\widehat{G}_{p,t+k|t} = k_t^* \, G_{clr,t+k|t}. \tag{10}$$

The persistence model is highly accurate for many parts of the year when there are cloudless days or even for highly overcast as long as the cloud cover conditions remain relatively unchanged. Because the persistence model incurs large errors only during highly variable days, the persistence errors directly help to describe the intrinsic random variability in solar irradiance (see (Hoff et al., 2010) for more on solar variability, and (Marquez and Coimbra, 2013b) for further discussion on bench-marking forecast quality with persistence). Because of this property, the persistence model is the best way to compare models developed and tested over different data sets (Marquez and Coimbra, 2013b).

Many clear-sky models can be found in literature (Ineichen, 2006) but in this work we developed a model specific to our data. Here, the clear-sky GHI is given by a polynomial expression that depends on the cosine of the solar zenith angle θ:

$$G_{clr,t} = \sum_{n=0}^{3} a_n (cos(\theta_t))^n, \tag{11}$$

where the coefficients a_n are derived through the application of Least-Squares-Method (LSM) fit for clear-sky GHI data selected manually. The LSM fit returned the following coefficients:

$$a_n = \{-0.0036, 0.56, 1.10, -0.64\}.$$

The last model shown in Table 2 is the Cloud Motion Forecast (CMF) model presented in the study in Hammer et al. (2003), where $k_{f,t+k|t}^*$ is obtained by forecasting the cloud cover through frozen cloud translation.

9. Forecast results

In this section we analyze the performance of the forecasting models. First we present a qualitative error analysis, followed by a quantitative error analysis where we compare the results of the proposed methodology with results by other methods proposed in the literature.

9.1. Qualitative error analysis

In order to analyze the performance of the forecasting models we plot pairs of forecast values vs. measured values of GHI in Figs. 10 and 11. The data used in these figures corresponds to testing data for the forecasting horizons of 30 and 90 min, respectively. These plots allow us to observe the deviation from the perfect forecast (the 1:1 line). The scatter plots for the 30 min forecasting horizon (Fig. 10) show that the majority of the persistence model predictions cluster tightly around the 1:1 line but it also incurs large errors evidenced by the many points dispersed away from the 1:1 line (mostly for irradiance above $0.5 \, \text{kW} \, \text{m}^{-2}$). The $\widehat{G}_{f1}$ model shows an overall narrower spread of pairs $(G_{t+k}, \widehat{G}_{f1,t+k|t})$ which indicates that the information about the could cover retrieved from the satellite images and included in the $\widehat{G}_{f1}$ model is essential to avoid the large errors observed in the persistence model. The lack of auto regressive terms in this model causes a small increase in the error for clear-sky conditions when compared to the persistence model, as illustrated by a higher deviation from the 1:1 line, especially noticeable for points near $1 \, \text{kW} \, \text{m}^{-2}$. The third model $\widehat{G}_{f2}$ shows the good features of the previous 2 models: tight clustering about the 1:1 line for clear-sky conditions and an overall narrower scatter. Fig. 11 shows similar features for the forecasting horizon of 90 min. However, for this case the differences between the models $\widehat{G}_{f1}$ and $\widehat{G}_{f2}$ are smaller (the R^2 is the same). This may be understood in view that the inclusion of lagged values of GHI in the model $\widehat{G}_{f2}$ becomes irrelevant as the correlation between G_t and G_{t+k} decreases with increasing k.

In a second type of qualitative analysis, the residuals for the forecasting models for the testing data are analyzed by means of the sample auto-correlation function (ACF). The
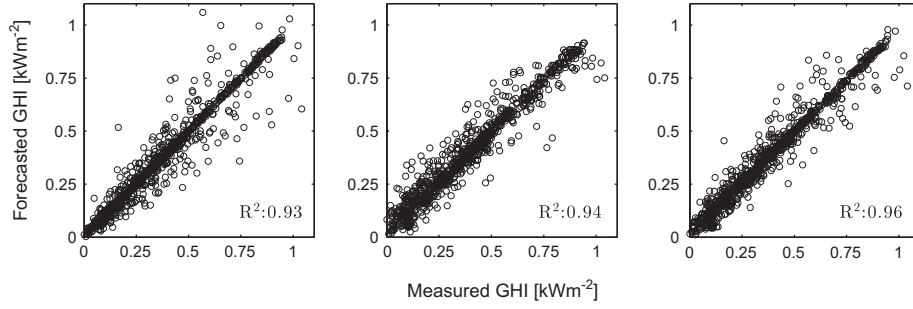
Fig. 10. Scatter plots of measured and forecasted GHI values $(G_{t+k}, \widehat{G}_{f,t+k|t})$ for the 30-min forecast horizon. *Left*: $\widehat{G}_p$, *center*: $\widehat{G}_{f1}$, *right*: $\widehat{G}_{f2}$. The data shown in these plots corresponds to the testing set.
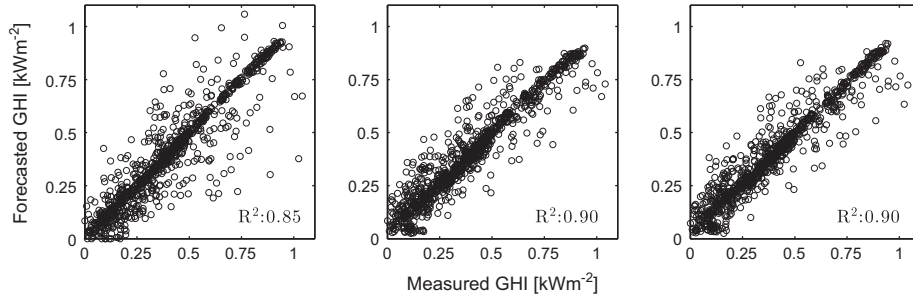

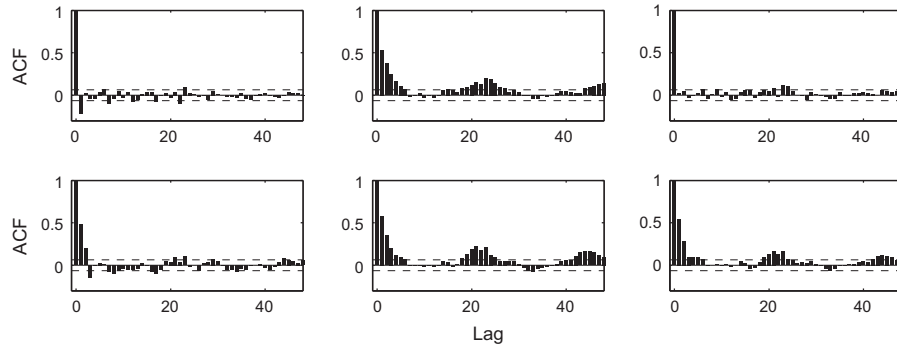Fig. 11. Same as the Fig. 10, but for the 90-min forecast horizon.

Fig. 12. Auto-correlation function for the residuals of the different forecasting models applied to the testing data. *Left*: $\widehat{G}_p$, *center*: $\widehat{G}_{f1}$, *right*: $\widehat{G}_{f2}$. *Top*: 30 min, *bottom*: 90 min.

ACF values for 48 lags in 30 min. steps are shown in Fig. 12 for the 30 and 90 min forecasting horizons. For the case of 30 min horizon both $\widehat{G}_p$ and $\widehat{G}_{f2}$ models show very small correlation values for all lags, which is representative of a white noise distribution of the residuals. On the other hand, the model $\widehat{G}_{f1}$ shows high ACF values for several lags. This is a clear sign that the error is not white noise and that the model could be further improved with the use of an auto-regressive model. For the time horizon of 90 min. all models show higher correlations for the first 2 lags.

9.2. Quantitative error analysis

The accuracy of the forecasting models is quantified using three evaluation criteria: the Mean Biased Error (MBE), the Root Mean Square Error (RMSE), the RMSE improvement. For a given forecasting horizon $t + k$ and for a given model f, MBE and RMSE are evaluated as:

$$\mathrm{MBE}_{f,k} = \frac{1}{N} \sum_{t \in DT} (G_{t+k} - \widehat{G}_{f,t+k|t}), \qquad (12)$$

and

$$\mathrm{RMSE}_{f,k} = \sqrt{\frac{1}{N} \sum_{t \in DT} (G_{t+k} - \widehat{G}_{f,t+k|t})^2}, \qquad (13)$$

where G_{t+k} is the measured value, $\widehat{G}_{f,t+k|t}$ is the forecasted value, and the summation is carried over $t \in DT$ day-time values only. The day-time sets satisfy the criteria, cos $(\theta_t) > 0.15$ which also excludes the early and late parts of

Table 3
The Mean Bias Error (MBE), the Root Mean Square Error (RMSE), and the RMSE improvements over the persistence model. See Table 2 for model descriptions. Values denoted in bold represent best performance for a given set and forecast horizon. RMSE improvements for the CMF model are estimated from Fig. 9 of (Hammer et al., 2003).

Forecast Horizon	Set	MBE (w/m²)			RMSE (w/m²)		1-RMSE/RMSE$_p$ (%)			CMF
		$\widehat{G}_p$	$\widehat{G}_{f1}$	$\widehat{G}_{f2}$	$\widehat{G}_p$	$\widehat{G}_{f1}$	$\widehat{G}_{f2}$	$\widehat{G}_{f1}$	$\widehat{G}_{f2}$	
30 min.	Train/Val.	2.01	1.83	**1.78**	63.27	56.79	**51.89**	10.25	**17.99**	8
	Test	**1.29**	−7.22	−2.07	61.24	57.84	**49.79**	5.56	**18.70**	
60 min.	Train/Val.	4.75	1.87	**0.64**	80.58	66.68	**62.02**	17.28	**23.04**	18
	Test	3.97	−7.06	−3.87	77.94	68.11	**62.42**	12.61	**19.91**	
90 min.	Train/Val.	10.23	**1.85**	2.49	92.88	74.70	**72.10**	19.58	**22.24**	17
	Test	8.63	−6.08	−3.28	93.52	75.31	**73.06**	19.47	**21.88**	
120 min.	Train/Val.	16.37	3.40	**2.95**	105.00	**79.32**	79.77	**24.46**	24.03	16
	Test	14.52	−2.26	**0.01**	107.47	**80.21**	81.26	**25.27**	24.39	

the day as these parts of the days would cause excessive difficulties in the image processing (see Section 3).

We apply the accuracy metrics independently to the two existent data sets–training/validation and testing. Where possible, we also include comparisons with existing satellite-based solar forecasts such as in Hammer et al. (2003); Perez et al. (2010b). The evaluations are performed for the forecasting horizons of 30, 60, 90, and 120 min for the 30-min averaged GHI values. The results from this evaluation are listed in Table 3. Additionally Table 3 also lists the RMSE improvement with respect to the persistence model:1-RMSE/RMSE$_p$, where RMSE$_p$ is the RMSE for the clear-sky-deviation persistence model.

The best RMSEs are generally obtained with the model $\widehat{G}_{f2}$ for all forecasting horizons except for 120 min. in which case $\widehat{G}_{f1}$ displays the lowest RMSE. As a measure of the forecasting skill, we consider the relative RMSE improvement over persistence (also shown in Table 3). The relative improvements range between 5% and 25% for the proposed forecast methods. The RMSE improvement with respect to the persistence model is considerably better as the forecasting horizon increases.

The RMSEs also lie in the range obtained in the satellite-based solar irradiance forecasts for a single location by other researches (Hammer et al., 1999a,a, 2001, 2003). In order to compare the forecast performance of Hammer et al. (2003), we have estimated the RMSE forecast skill values from Fig. 9 of Hammer et al. (2003), which are shown in the last column of Table 3. From these observations, it appears that the proposed methodology compares favorably with existing satellite-based forecasting models.

The forecasting methodology proposed here can also be compared with the Cloud Motion Forecasting (CMF) results obtained in Perez et al. (2010a) by comparing the relative improvements of RMSEs over the persistence model. RMSE data obtained from Perez et al. (2010a) is plotted in Fig. 13. The plots show the RMSEs of cloud motion-based hourly GHI forecasts and persistence RMSEs for forecasting horizons of 1–6 h. There are two similarities to point out between the results in Perez et al. (2010a) and the results obtained here: the first observation
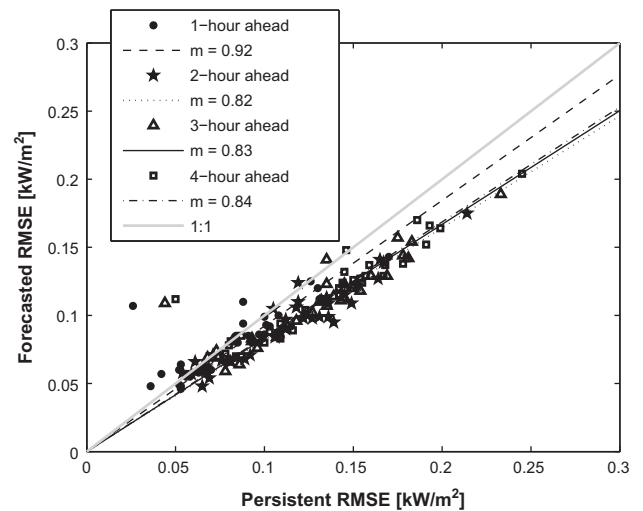

Fig. 13. RMSE of forecast and persistence models for forecasting horizons 1–4 h ahead. Data obtained from (Perez et al., 2010a). The different slopes m are used to estimate the relative forecast improvements over persistence: 1-RMSE$_f$/RMSE$_p$ = 1 − m.

is that the smallest relative improvements are obtained for the first forecasting horizon, as can be seen by estimating the relative improvements as given by the slopes ($1 − \text{RMSE}_f/\text{RMSE}_p = 1 − m$). For the 30-min ahead forecasting of the 30-min average GHI, we obtain an improvement of $\approx 5 − 19\%$, whereas (Perez et al., 2010a), reports an improvement of $\approx 10\%$. These forecasting horizons correspond to 1 time-step (Δt) forecasting horizon for the respective studies. The second observation is that, for multiple (k) time-step ($k\Delta t$) forecasting horizons, the relative improvements are consistently $\approx 20\%$.

From the values in Table 3 we can also observe that the beneficial effect of including GHI as an input begins to decrease with longer time horizons, to the point that for horizons of 120 min, the best model is $\widehat{G}_{f1}$. Therefore, at longer forecast time horizons the accuracy seems to depend more on the satellite derived information rather than on the solar irradiance. This is expected given the decrease in correlation between G_t and G_{t+k} with increasing k.

A final observation is that the similarity between the results for the training/validation and testing data sets indicate that no significant overtraining (overfitting) is present, and that the forecast evaluation procedure is robust.

10. Conclusions

This work presents a forecasting approach for predicting global horizontal solar irradiance (GHI) based on satellite observations and Artificial Neural Networks (ANNs). The satellite images are processed using cloud indexing and velocimetry algorithms so that cloud fraction variables can be used as inputs to the ANN model. The forecasts cover horizons of 30, 60, 90, and 120 min for 30-min averaged sets of GHI data (time compression ratios varying from 1 to 4). The resulting forecast root-mean-squared errors (RMSEs) and relative improvements over the persistence model indicates that the proposed forecast method performs similarly or improves upon results obtained in other studies (Hammer et al., 1999b, 2001; Perez et al., 2010a). The improvements over persistence, in particular, provide a very reliable way of comparing forecast models. The improvements are ≈5–19% for 1 time-step forecasts (compression ratio 1), and ≈10–25% for multiple time-step forecasts (compression ratios > 1).

The novelty of this study is the hybridization of detailed satellite-image processing as inputs for an ANN-based forecast method. Although these models are at least on par with conventional frozen cloud translation approaches they used rather simple image processing to identify clouds (for instance we cannot distinguish snow in the ground from clouds). We anticipate that more sophisticated stochastic learning approaches will systematically decrease translation errors from the image processing phase, which in turn will lead to greater accuracy and reliability of solar forecasting. For example, classification techniques can be used to identify snow, fog, cloud-type, cloud texture, etc. and include these classes as part of the ANN optimization phase. Also, given that the main goal in this work was to demonstrate the hybridization of satellite-images and ANNs we devoted less effort to the optimization of the ANN architecture, the selection of lagged inputs, the pre-processing of the input variables. Pedro and Coimbra (2012) show that more elaborated pre-processing combined with topology optimization of the ANNs can improve substantially the forecasting performance. This limitation in the proposed method can easily be resolved (albeit at a significant computational cost) by integrating the method with an optimization algorithm (Pedro and Coimbra, 2012).

Finally, this work indicates that hybrid methods that employ a combination of stochastic learning, ground and/or remote sensing image processing, and ground telemetry do offer substantial advantages in accuracy and robustness for the development of high-fidelity solar forecasts.

References

Bacher, P., Madsen, H., Nielsen, H.A., 2009. Online short-term solar power forecasting. Solar Energy 83, 1772–1783.

Bishop, C., 1994. Neural networks and their applications. Review of Scientific Instruments 65, 1803–1832.

California Independent System Operator (CAISO), 2010. Integration of renewable resources: operational requirements and generation fleet capability at 20 percent RPS. <http://www.caiso.com/2804/2804d036401f0.pdf>

Cano, D., Monget, J., Albuisson, M., Guillard, H., Regas, N., Wald, L., 1986. A method for the determination of the global solar-radiation from meteorological satellite data. Solar Energy 37 (1), 31–39.

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

Hammer, A., Heinemann, C., Hoyer, C., Lorenz, E., 2001. Satellite based short-term forecasting of solar irradiance–comparison of methods and error analysis. In: The 2001 EUMETSAT meteorological satellite data user's conference, pp. 677–684.

Hammer, A., Heinemann, C., Lorenz, E., Lückeke, B., 1999a. Short-term forecasting of solar radiation based on image analysis of meteosat data. In: TEUMETSAT meteorological satellite data user's conference. pp. 331–337.

Hammer, A., Heinemann, D., Hoyer, C., Kuhlemann, R., Lorenz, E., Müller, R., G., B.H., 2003. Solar energy assessment using remote sensing technologies. Remote Sensing of Environment 86, 423–432.

Hammer, A., Heinemann, D., Lorenz, E., Ckehe, B.L., 1999b. Short-term forecasting of solar radiation: a statistical approach using satellite data. Solar Energy 67, 139–150.

Hoff, T.E., Perez, R., 2010. Quantifying pv power output variability. Solar Energy 84 (10), 1782–1793.

Ineichen, P., 2006. Comparison of eight clear sky broadband models against 16 independent data banks. Solar Energy 80, 468–478.

Kostylev, V., Pavlovski, A., 2011. Solar power forecasting performance towards industry standards. In: 92nd American Meteorological Society Annual Meeting, pp. 1–8.

Lew, D., Piwko, R., 2010. Western wind and solar integration study. Tech. rep., National Renewable Energy Laboratories.

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., 2013a. Intra-hour DNI forecasting methodology based on cloud tracking image analysis. Solar Energy 91, 327–336.

Marquez, R., Coimbra, C.F.M., 2013b. Proposed metric for evaluating solar forecasting models. ASME Journal of Solar Energy Engineering 135, 0110161–0110169.

Marquez, R., Gueorguiev, V.G., Coimbra, C.F.M., 2013. Forecasting of global horizontal irradiance using sky cover indices. ASME Journal of Solar Energy Engineering 135, 0110171–0110175.

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.

Mellit, A., 2008. Artificial Intelligence technique for modelling and forecasting of solar radiation data: a review. International Journal of Artificial Intelligence and Soft Computing 1, 52–76.

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.oceanwave.jp/softwares/mpiv>.

Pedro, H.T.C., Coimbra, C.F.M., 2012. Assessment of forecasting techniques for solar power output with no exogenous inputs. Solar Energy 86 (7), 2017–2028.

Perez, R., Kivalov, S., J., S., Hemker, K.J., Renne, D., Hoff, T.E., 2010a. Validation of short and medium term operational solar radiation forecasts. Solar Energy 84, 2161–2172.

Perez, R., Kivalov, S., Schlemmer, J., Hemker, K., Renne, D., Hoff, T.E., 2010b. Validation of short and medium term operational solar radiation forecasts in the us. Solar Energy 84 (5), 2161–2172.

Reikard, G., 2009. Predicting solar radiation at high resolutions: a comparison of time series forecasts. Solar Energy 83 (3), 342–349.

Schillings, C., Mannstein, H., Meyer, R., 2004. Operational method for deriving high resolution direct normal irradiance from satellite data. Solar Energy 76, 475–484.