Ricardo Marquez
e-mail: rmarquez3@ucmerced.edu

Vesselin G. Gueorguiev
e-mail: vgueorguiev@ucmerced.edu

Mechanical Engineering and
Applied Mechanics (MEAM),
School of Engineering,
University of California,
Merced, CA 95343

Carlos F. M. Coimbra[1]
Mem. ASME
Mechanical and Aerospace Engineering (MAE),
Jacobs School of Engineering,
University of California,
San Diego, CA 92093
e-mail: ccoimbra@ucsd.edu

Forecasting of Global Horizontal Irradiance Using Sky Cover Indices

This work discusses the relevance of three sky cover (SC) indices for solar radiation modeling and forecasting. The three indices are global in the sense that they integrate relevant information from the whole sky and thus encode cloud cover information. However, the three indices also emphasize different specific meteorological processes and sky radiosity components. The three indices are derived from the observed cloud cover via total sky imager (TSI), via measurements of the infrared radiation (IR), and via pyranometer measurements of global horizontal irradiance (GHI). We enhance the correlations between these three indices by choosing optimal expressions that are benchmarked against the GHI SC index. The similarity of the three indices allows for a good qualitative approximation of the GHI irradiance when using any of the other two indices. An artificial neural network (ANN) algorithm is employed to improve the quantitative modeling of the GHI sky cover index, thus improving significantly the forecasting details of GHI when all three indices are used. [DOI: 10.1115/1.4007497]

1 Introduction

In the context of the near-future scenario where a significant amount of variable energy resources is used to produce electricity, power producers, utility companies, and independent system operators will rely heavily on streaming information to balance and manage the variability of energy supply and demand. SC is the most critical variable in forecasting solar energy power generation. Satellite images are common source of cloud cover information but they lack the detailed temporal and spatial resolution required to accurately monitor the operational effects on solar energy for a given site. A shortcoming of using satellite images for cloud information is the lack of detailed temporal and spatial resolution required to accurately monitor the operational effects on solar energy for a given site [1]. An alternative approach is to use ground based sky images from TSI where data can be gathered at high frequency, but even then such information is only available during daytime. Thus, night cloud cover information that is relevant for forecasting solar power generation in the early morning hours is not registered. An additional problem of completely relying on TSI data is the high costs associated with imaging instruments, which limits their utilization for covering large regions [2,3].

One of the objectives of this work is to address some of these shortfalls by developing a technique for obtaining cloud cover information using infrared radiometers (IR) as a complement to the development of cloud cover methods based on the visible part of the electromagnetic radiation spectra. Some previous works have examined how to extrapolate cloud information from solar radiation (SR) and other recent works have also included infrared radiation to predict cloud cover [3,4]. Measurements that use IR are cheaper relative to sky imaging measurements, are characterized by high measurement and collection frequency, and are already located at many standard weather monitoring stations [2], and therefore wider coverage is possible. Another advantage of using IR radiometers is that they are sensitive to cloud cover during the night time [4], which is critical to forecast morning shoulder ramping rates.

The intended application of our results is to incorporate the IR measurements as inputs to solar radiation forecast models. To this extent, solar forecasting methodologies rely on cloud cover information with varying rates of success [1,5,6]. In Refs. [1] and [5], sky cover is obtained from forecasts provided by the National Weather Service (NWS) where the data sets can be used to produce same-day and several days ahead hourly solar forecasts. Cao and Lin [6] used cloud cover data retrieved from a weather station forecasts given in specific category format: overcast, sunny, cloudy, and cloudy and sunny where the categories were later processed using defuzzification techniques. These previous approaches, however, have generally neglected the use of infrared radiation and thus ignored opportunities for improvements in forecasting performance.

Our goal is to correlate TSI cloud cover measurements, IR, and SR at the Earth's surface and to use this information as input for modeling and forecasting of the SR. In order to develop the various correlations, we gather TSI and IR measurements and correlate both measurements with SR data. We note that each of these measurements is affected by the amount of the cloud cover present. There are inherited limitations to developing such correlations since they depict different information about the underlining atmospheric processes. For example, sky cover values obtained from the TSI are typically based on the proportion of pixel values above some threshold in the visible range. The IR measurements are sensitive to many factors, including surrounding temperature, sky temperature, relative humidity, atmospheric composition, and the presence of nearby clouds. Out of all these factors, cloud presence contributes the most to the variations in IR and our objective is to isolate these effects. The variability of SR is also strongly dependent on clouds.

Our approach correlates cloud effects on TSI and IR with the cloud effects on GHI. We then produce estimates of GHI based on TSI and IR. Section 2 presents the data sets used and explains the methodology of our analysis. In Sec. 3, we study three SC indices based on IR, TSI, and GHI measurements. Section 4 shows the results of the GHI estimates based on the IR and TSI indices. Section 5 demonstrates the importance of the IR and TSI information for the improvement of GHI forecasts based on ANN algorithms. Concluding remarks are given in Sec. 6.

2 Data Collection and Methodology

The data used in this work correspond to the period between October 3 and December 3, 2010 at the University of California

[1]Corresponding author.

Contributed by the Solar Energy Division of ASME for publication in the JOURNAL OF SOLAR ENERGY ENGINEERING. Manuscript received April 26, 2011; final manuscript received December 23, 2011; published online October 23, 2012. Assoc. Editor: Carsten Hoyer-Klick.

Merced Solar Observatory. This period is characterized by a mixture of clear and overcast days. All instruments are located within 4 m radius of each other and are strategically placed so that shading and interference effects are minimized. Additional meteorological data sets, such as dewpoint temperature and soil temperature, are taken from the California Irrigation Management and Information Systems (CIMIS) at the Merced weather station (#148), which is located within 7 km of the UC Merced solar observatory (sol.ucmerced.edu).

2.1 Global Horizontal Irradiance. We consider the effects of cloud cover on GHI. An Eppley PSP (precision spectral pyranometer) measures irradiance at wavelengths between 0.285 μm and 2.80 μm, which comprises approximately 98.5% the solar spectrum. The amount of GHI has a diurnal dependence on the solar zenith angle. In order to remove this dependence, to isolate the cloud effects on GHI, and to compare to sky cover indices based on the TSI and IR measured values, we consider the clearness index K_g

$$K_g = \mathrm{GHI}/\mathrm{GHI}_{\mathrm{clear}} \tag{1}$$

$\mathrm{GHI}_{\mathrm{clear}}$ is computed using the ESRA model adopted from Refs. [7] and [8]. In order to choose an appropriate value for the Linke-turbidity factor (T_L), the following clear-days from the 2010 data set were used: October 10–15, and 20, and November 1, 15, and 24. By varying the T_L and comparing with the clear-days, a suitable value for T_L was found to be 7. On the clear-sky data set, the model produces an MBE and RMSE values of $-17.1\,\mathrm{W/m^2}$ and $21.0\,\mathrm{W/m^2}$, respectively. These errors are small compared to the larger GHI fluctuations resulting from cloud effects, so we do not expect our $\mathrm{GHI}_{\mathrm{clear}}$ estimates to contribute much to the random and systematic errors on the subsequent correlations.

From the above equation, $K_g = 1$ for clear days and a very small value for cloudy days depending on total cloud cover. Sky cover variability based on GHI can be defined as

$$\mathrm{SC}_{\mathrm{GHI}} = 1 - \mathrm{GHI}/\mathrm{GHI}_{\mathrm{clear}} = 1 - K_g \tag{2}$$

These sky cover estimates are compared in the next sections. In terms of solar modeling and forecasting, the most important sky cover quantity is the one based on the clearness index (K_g). Therefore, the TSI- and IR-based cloud cover quantities are benchmarked according to the clearness index values.

2.2 Infrared Measurements. The instruments used for measuring radiation are oriented on horizontal surfaces. The solar irradiance was measured with a pyranometer (PSP), and the infrared radiation was measured with a pyrgeometer (PIR), both of which were manufactured by Eppley Labs [9,10]. In the PIR, the case and dome temperatures are recorded along with net infrared radiation (c). The IR is calculated according to [11]

$$\mathrm{IR} = \mathrm{IR}_{\mathrm{net}} + \sigma T_{\mathrm{c}}^4 - k\sigma(T_{\mathrm{d}}^4 - T_{\mathrm{c}}^4) \tag{3}$$

where T_{d} is the instrument dome temperature, T_{c} is the instrument case temperature, k is constant with value $k = 3.5$ as recommended by the manufacturer, and σ is the Stephan–Boltzmann constant. The detector which measures K_g is sensitive to radiation at wavelengths between 3.5 μm and 50 μm. In this range of wavelengths, radiation is mostly emitted from objects close to 300 K and virtually no radiation from the Sun is received at these wavelengths. The intensity of the infrared radiation from the Sun is about two orders of magnitude smaller than the intensity in the visible spectrum. It is also about an order of magnitude smaller than the intensity of the IR radiation emitted by the Earth. In this respect, the clear-sky values of IR are approximated by the Stephan–Boltzmann relationship with emittance ε that depends on the dewpoint temperature (T_{dewpoint}) and time of day [12]

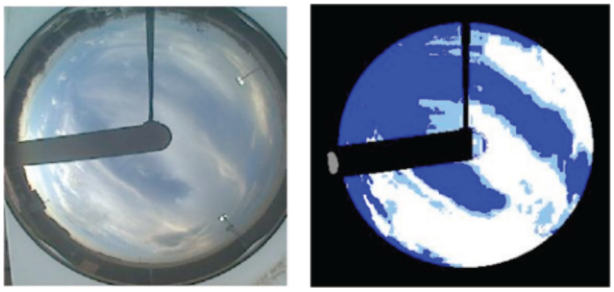

Fig. 1 Left: example of image taken by the TSI for a cloudy day. Right: processed image using manufacturer's image classification algorithms. White regions on lower and top right region indicate opaque clouds while thin clouds are represented by intermediate shading.

$$\mathrm{IR}_{\mathrm{clear}} = \varepsilon\sigma T_{\mathrm{ambient}}^4 \tag{4}$$

$$\varepsilon = 06921 + 000244T_{\mathrm{dewpoint}} + 005101\sin(15\mathrm{t}) \tag{5}$$

where t is the number of hours from midnight and the ambient temperature is used in place of the sky temperature. The time variable (t) helps to remove the ambient temperature diurnal patterns introduced by estimating $\mathrm{IR}_{\mathrm{clear}}$ from T_{ambient}. A more reasonable way to remove the diurnal patterns is to model ε as a function of relative humidity. Because the above equations give satisfactory estimates of $\mathrm{IR}_{\mathrm{clear}}$, we defer the aforementioned corrections.

2.3 Total Sky Imager Data. Images of the sky are taken by the Yankee Environmental Systems, Inc. TSI instrument [13]. The images are processed with the manufacturer's software that classifies pixel values for images taken with a camera located above a rotating hemispherical mirror. The mirror contains a band that continuously blocks the mirror from the sun (see Fig. 1). The processed sky cover information from the TSI includes fractional sky cover due to thin and opaque clouds, which are distinguished by the light blue and white shaded areas in Fig. 1. The proportion of these areas to the total circular area is classified as fractional sky cover. Images are taken and processed at 1 s intervals from which hourly values are subsequently calculated.

The TSI does not produce other cloud properties such as cloud type, speed, or height. The information in the above list is given as a time-series which only includes day-time values since the sky-imager does not process images in the dark. As shown in Fig. 1, the built-in classification algorithms do not always classify all the clouds preset. Developing more accurate classifications is essential for a more precise use of the TSI. However, we do not address this issue here. Values calculated with the YES built-in algorithms are used in this paper.

3 Correlating Sky Cover Indices

Cloud cover information is based on TSI, IR, and GHI measured values. The relations given in Table 1 show how SC is computed using data from each of the instruments. Note that the SC

Table 1 Expressions defining sky cover (SC) indices derived from GHI, TSI, and IR data. The subscript "clear" indicates data values associated with clear sky days.

Data type	Sky cover index
Visible radiation	$\mathrm{SC}_{\mathrm{GHI}} = 1 - \mathrm{GHI}/\mathrm{GHI}_{\mathrm{clear}}$
Total sky imager	$\mathrm{SC}_{\mathrm{TSI}} = \mathrm{aSC}_{\mathrm{thin}} + \mathrm{bSC}_{\mathrm{opaque}} + c$
Infrared radiation	$\mathrm{SC}_{\mathrm{IR}} = c_2\delta_{\mathrm{IR}}^2 + c_1\delta_{\mathrm{IR}} + c_0$, $\delta_{\mathrm{IR}} = \mathrm{IR} - \mathrm{IR}_{\mathrm{clear}}$

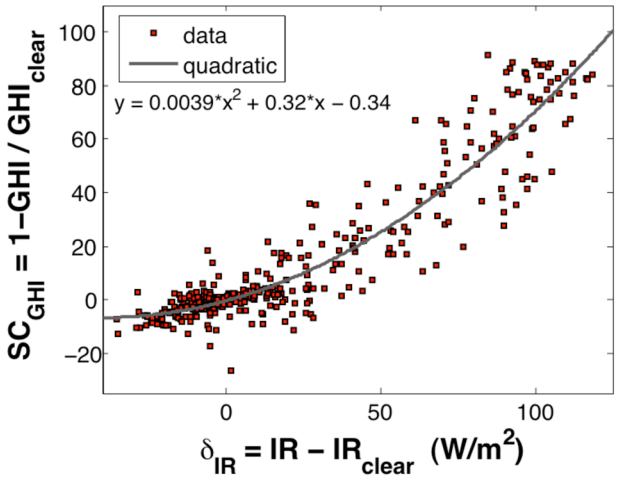

Fig. 2 Scatter plots of SC_{GHI} versus δ_{IR}. The line is a quadratic fit used to define SC_{IR}. The parameters of the polynomial are also shown under the legend box.

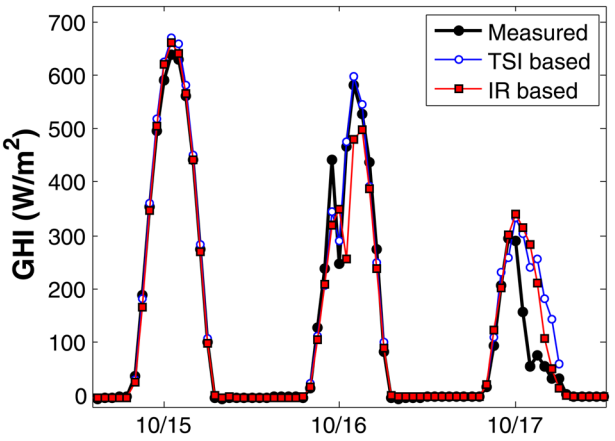

Fig. 4 GHI time series measured and estimated based on SC_{IR} and SC_{TSI}. Night values are removed from this time-series plot.

quantities based on the TSI and IR contain parameters that are used for fitting the data on SC that is based on GHI. This is because in this study we are specifically interested in extracting predictive information to forecast GHI. For the case of the TSI, we assume that a linear combination of the thin and opaque sky cover correlates well with the SC_{GHI}. This approach is more general than simply adding thin and opaque values because the coefficients may help to correct some of the errors introduced during image processing of the sky images. For IR, we base SC_{IR} on the difference between measured IR and estimated clear-sky values of IR (IR_{clear}). The differences are denoted as δ_{IR}, which we estimate (based on Fig. 2) can be fitted quadratically with SC_{GHI}.

In order to apply the correlations properly, it is necessary to remove early morning and late evening hours when estimating the parameters. The reason for this is because errors are easily encountered when calculating SC_{GHI} when GHI_{clear} is small. Also, for a few days in the data set, condensation on the domes of the PSP instruments occurred due to morning fog. Thus, to estimate the parameters for SC_{TSI} and SC_{IR} we use only the day-time data set from 10:00 a.m. to 4:00 p.m. to avoid data contamination.

Figure 2 clearly shows strong correlation effects among the GHI and IR sky cover indices. In particular, it seems appropriate to estimate the cloud effects as a function of the difference $\delta_{IR} = (IR\text{-}IR_{clear})$ using a quadratic polynomial fit (see Table 1). The polynomial coefficients are calculated based on a least squares regression fit as shown in the figure. As for the GHI_{clear}, we used the set of 10 clear days. The fit produces an $R^2 = 0.884$ and RMSE = 9.46 W/m^2.

We obtain the values of the parameters in Table 1 for the expression of SC_{TSI} to be $a = -0.116$, $b = 1.31$, and $c = 0.091$ by benchmarking the SC_{TSI} with SC_{GHI}. This fit produces an $R^2 = 0.823$ and RMSE = 11.8 W/m^2, which is comparable to the values obtained from correlating SC_{IR} with SC_{GHI}.

The level of quantitative and qualitative agreement of the three sky cover indices discussed above is shown in Fig. 3.

4 Calculating GHI From Sky Cover Indices

In Sec. 3, we derived relationships between sky cover quantities inferred from TSI, IR, and GHI measurements. Polynomial fits were used to correlate TSI and IR sky cover with GHI sky cover. From these fits, we can approximate GHI from TSI and IR by manipulating the equations and using the curve fits for each SC

$$GHI_{estimated,IR} = (1 - SC_{IR})GHI_{clear} \quad (6)$$

$$GHI_{estimated,TSI} = (1 - SC_{TSI})GHI_{clear} \quad (7)$$

Applying the above GHI estimates from IR and TSI measured values, the time-series in Fig. 4 is obtained for October 15–17. Note that night values are automatically set to zero due to the multiplication of GHI_{clear}. Both of the estimates agree very well for the days shown, particularly the clearest day. Note the locations of the cusps especially for November 16. The close agreements obtained here are an indication that the formulation prescribed so far for relating the values obtained from the GHI, IR, and TSI instrumentation is correct, and thus motivates the use of the SC indices as valid inputs for forecasting models.

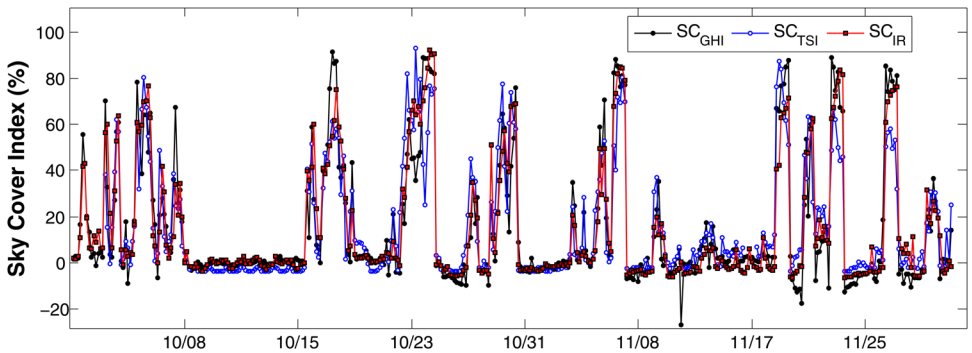

Fig. 3 Time series of the sky cover index inferred from the TSI, IR, and GHI for October and November 2010. The night, dawn, and dusk values are excluded from the graph. The instrument maintenance day (Nov. 17) was also removed from the set shown.

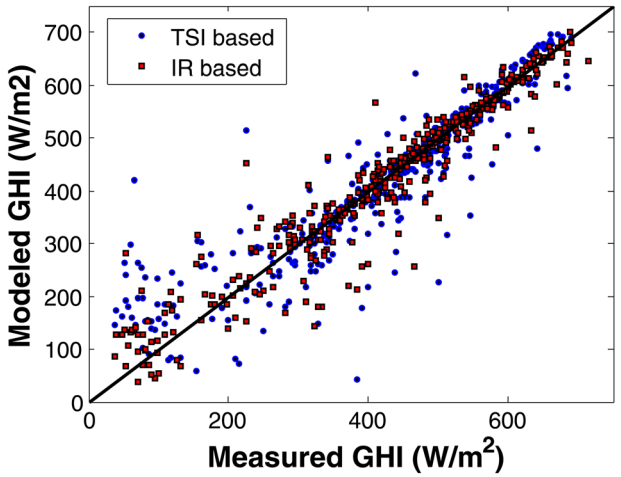

Fig. 5 Scatter plot for comparing GHI based on IR and TSI. The correlation coefficients (R^2) for IR and TSI estimates are 0.935 and 0.923, respectively.

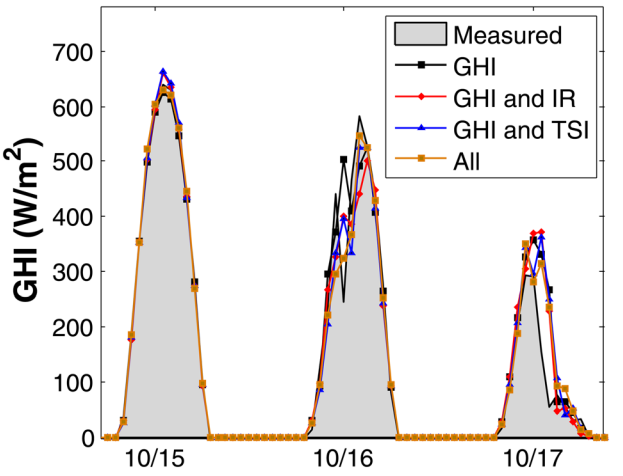

Fig. 6 GHI time series for measured and forecasted values based on SC_{GHI}, SC_{IR}, and SC_{TSI}. The legend shows which inputs were included in the forecasting model. Night values are not included.

A scatter plot (see Fig. 5) of the estimated GHI values versus the measured values shows that the GHI estimates are very reliable as the squared correlation coefficients indicate, e.g., $R^2 = 0.935$ for the IR-based model and $R^2 = 0.923$ for the TSI-based model. In terms of modeling GHI, TSI and IR produce similar results with the spread of the IR estimates showing slightly better results.

In summary, the following simple algorithm can be used for modeling GHI based on TSI and IR measured values:

1. calculate IR_{clear} from temperature and dewpoint temperature measurements
2. taking the difference between IR and IR_{clear} compute SC_{IR} using the quadratic polynomial
3. from SC_{IR} estimate GHI (see Eq. (6))

Alternatively, one can use a linear combination of opaque and thin TSI data to compute SC_{TSI} and then from SC_{TSI} estimate GHI (see Eq. (7)).

5 ANN Forecasting of Solar Radiation Using Sky Cover Indices

In this section, we apply the results obtained in the previous sections to the development of a forecast model for GHI. To the knowledge of the authors, no forecast methodology based on TSI based sky cover (cloud effects) for hourly and intra-hourly prediction have demonstrated major gains in accuracy, and none have used IR. The approach we adopt is to forecast the GHI-based SC index at 1 h ahead intervals, then use Eq. (2) to revert back to the GHI value.

The forecasting model is mathematically represented as

$$SC_{GHI}(t+1) = f(SC_{GHI}(t),\quad SC_{GHI}(t-1),\ldots,SC_{GHI}(t-n),\\ SC_{IR}(t),\quad SC_{IR}(t-1),\ldots,SC_{IR}(t-n),\\ SC_{TSI}(t),\quad SC_{TSI}(t-1),\ldots,SC_{TSI}(t-n)) \tag{8}$$

where the forecasting input variables include previous values of the SC indices at times t, $t-1$ h, and up to $t-n$ h. The model $f()$ can be a function based on a number of statistical forecast modeling approaches including ones used in solar irradiance modeling (ARIMA, ANNs, k-NN) [14,15]. We use ANNs as these models have a proven record for solar irradiance forecasting [15].

Table 2 Statistical metrics (R^2, root-mean-squared deviation, and mean-bias error) for the ANN forecast results based on SC_{GHI}, SC_{IR}, SC_{TSI}, and permutations. Night values were ignored in calculations. Bold values indicate top performance.

Inputs			Error metrics		
SC_{GHI}	SC_{IR}	SC_{TSI}	R^2	RMSE (W/m^2)	MBE (W/m^2)
√	—	—	0.940	50.2	**1.6**
√	√	—	0.967	37.2	**1.6**
√	—	√	0.949	46.6	1.7
√	√	√	**0.970**	**35.9**	1.7

The ANN model that we select is a feed-forward type composed of one input layer, one hidden layer, and one output layer. The input layer is comprised the inputs according to the arguments of Eq. (8) and the output layer is the value that we would like to predict SC_{GHI} $(t+1)$. The number of time delays is set to $n = 3$. The number of neurons in the hidden layer is fixed at 10. The activation functions of the hidden layers are sigmoidal and the output neuron has a linear activation function. We use Bayesian regularization to adjust the weights of the ANN synapses, and the models are trained until convergence of the weight values [16].

In order to identify which time-series variables contribute to the best forecasting performance, we apply the ANN model using four input combinations. The first input combination uses only the GHI-based SC index time-series, the second input set consists of SC indices based on GHI and IR, the third input set consists of SC indices based on GHI and TSI, and the fourth input set consists of all the SC indices.

Figure 6 shows the results of the ANN forecasting using various SC inputs. The days shown are the same as the ones in Fig. 4. One can immediately see the improvement over the results depicted in Fig. 4. In particular, the position of the cusp on October 16 is properly reproduced by all ANN that use an additional SC input besides the GHI data. By looking at October 17 one concludes that the ANN forecasting that uses all three SC indices seems is more accurate. These observations are yet to be confirmed by examining more cloudy days. However, as seen from Table 2, the advantage of using TSI and IR data together is clear.

6 Conclusions

We examined different strategies for inferring cloud cover from visible image processing and IR measurements as they pertain to the characterization of solar radiation at the ground level. We show that, by defining pertinent sky cover indices, it is possible to accurately model GHI based on measured values from the TSI and IR data. This work suggests that predictive information is embedded in the measurements obtained from the TSI and IR instrumentation, and that this information should be considered as relevant inputs for solar forecast models. A simple algorithm was developed and tested over several clear and cloudy days in October and November 2010 at the University of California Merced site in order to estimate GHI from both TSI and IR hourly values. Further improvement on GHI modeling is achieved by considering artificial neural network forecast models using various combinations of sky cover indices.

We identify two distinctive sources of error that can be improved regarding predictive quality for solar modeling and forecasting include (1) improved understanding of IR_{clear}, and (2) improved TSI image processing and classification algorithms. Because currently there has been only modest success in including the effects of cloud cover as obtained from meteorological station predictions [1,5,6], we propose that the methodology suggested in the present work shows promise in generating improved intra-hour solar forecasting models.

Achnowledgment

The authors gratefully acknowledge the partial financial support given by the California Energy Commission (CEC) under the PIER RESCO Project PIR-07-036, by the National Science Foundation (NSF) CNS division Grant No. 0923586, and also by the Eugene Cotta-Robles (ECR) Fellowship program of the University of California and the Southern California Edison Fellowship to UC Merced. Seed and continued support by the Center for Information Technology Research in the Interest of Society (CITRIS) is gratefully appreciated.

References

[1] Marquez, R., and Coimbra, C. F. M., 2011, "Forecasting of Global and Direct Solar Irradiance Using Stochastic Learning Methods, Ground Experiments and the NWS Database," Sol. Energy, **85**(5), pp. 746–756.
[2] Long, C. N., Barnard, J. C., Calbo, J., and Ackerman, T. P., 2005, "Using Integrated Surface Radiation Measurements to Infer Cloud Properties," Ninth Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface, http://ams.confex.com/ams/Annual2005/techprogram/paper_83452.htm
[3] Long, C. N., Ackerman, T. P., Gaustad, K. L., and Cole, J. N. S., 2006, "Estimation of Fractional Sky Cover From Broadband Shortwave Radiometer Measurements," J. Geophys. Res., **111**(D11204), pp. 1–11.
[4] Thurairajah, B., and Shaw, J. A., 2005, "Cloud Statistics Measured With the Infrared Cloud Imager (ICI)," IEEE Trans. Geosci. Remote Sens., **43**(9), pp. 2000–2007.
[5] Perez, R., Kivalov, S., Schlemmer, J., Hemker, K., Renne, D., and Hoff, T. E., 2010, "Validation of Short and Medium Term Operational Solar Radiation Forecasts in the US," Sol. Energy, **84**(5), pp. 2161–2172.
[6] Cao, J., and Lin, X., 2008, "Study of Hourly and Daily Solar Irradiation Forecast Using Diagonal Recurrent Wavelet Neural Networks," Energy Convers. Manage., **49**(6), pp. 1396–1406.
[7] Zarzalejo, L. F., Polo, J., and Ramirez, L., 2004, Gc_model5_irradiance, Matlab computer program (CDROM) accompanying Ref. [8].
[8] Badescu, V., 2008, *Modeling Solar Radiation at the Earth Surface*, Springer-Verlag, Berlin.
[9] The Eppley Laboratory, Inc., 2011, *Precision Spectral Pyranometer*, The Eppley Laboratory, Inc., Newport, RI, http://eppleylab.com
[10] The Eppley Laboratory, Inc., 2011, *Precision Infrared Radiometer*, The Eppley Laboratory, Inc., Newport, RI, http://eppleylab.com
[11] Albrecht, B., and Cox, S. K., 1977, "Procedures for Improving Pyrgeometer Performance," J. Appl. Meteorol., **16**, pp. 188–197.
[12] Berdahl, P., and Fromberg, R., 1982, "The Thermal Radiance of Clear Skies," Sol. Energy, **29**, pp. 299–314.
[13] Yankee Environmental Systems, Inc., 2007, *TSI-880 Automatic Total Sky Imager*, Yankee Environmental Systems, Inc., Turners Falls, MA, http://www.yesinc.com
[14] Paoli, C., Voyant, C., Muselli, M., and Nivet, M., 2010, "Forecasting of Preprocessed Daily Solar Radiation Time Series Using Neural Networks," Sol. Energy, **84**(12), pp. 2146–2160.
[15] Mellit, A., 2008, "Artificial Intelligence Technique for Modelling and Forecasting of Solar Radiation Data: A Review," Int. J. Artif. Intell. Soft Comput., **1**, pp. 52–76.
[16] Bishop, C. M., 1995, *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford, UK.