Available online at www.sciencedirect.com

SciVerse ScienceDirect

Solar Energy 86 (2012) 3531–3543

SOLAR ENERGY

www.elsevier.com/locate/solener

Hourly solar irradiance time series forecasting using cloud cover index

Yang Dazhi [a,*], Panida Jirutitijaroen [a], Wilfred M. Walsh [b]

[a] Department of Electrical and Computer Engineering, National University of Singapore, Singapore
[b] Solar Energy Research Institute of Singapore, National University of Singapore, Singapore

Available online 14 September 2012

Communicated by: Associate Editor Christian Gueymard

Abstract

We apply time series analysis to forecast next hour solar irradiance including cloud cover effects. Three forecasting methods are proposed using different types of meteorological data as input parameters, namely, global horizontal irradiance (GHI), diffuse horizontal irradiance (DHI), direct normal irradiance (DNI) and cloud cover. The first method directly uses GHI to forecast next hour GHI through additive seasonal decomposition followed by an Auto-Regressive Integrated Moving Average (ARIMA) model. The second method forecasts DHI and DNI separately using additive seasonal decomposition followed by an ARIMA model and then combines the two forecasts to predict GHI using an atmospheric model. The third method considers cloud cover effects. An ARIMA model is used to predict cloud transients. GHI at different zenith angles and under different cloud cover conditions is constructed using nonlinear regression, i.e., we create a look-up table of GHI regression models for different cloud cover conditions. All three methods are tested using data from two weather stations in the USA: Miami and Orlando. It is found that forecasts using cloud cover information can improve the forecast accuracy.

Keywords: Cloud cover; Time series; Regression; ARIMA

1. Introduction

As solar photovoltaic electricity (PV) approaches price parity with grid electricity, the installed base of solar electricity systems rapidly increases. This creates opportunities for companies and challenges for electricity grid operators. Both groups require detailed, accurate information about how much photovoltaic electricity can be generated, where it will be generated and when it will and will not be generated.

The major drawback of solar energy lies in its non-continuous generation: PV power drops as the solar irradiance drops. Typically, cloud cover will cause many rapid changes in the irradiance during the day. A significant decrease in the grid's PV power delivery due to reduced irradiance can pose problems for grid operators who must compensate for the shortfall. Therefore, accurate short-term solar irradiance forecasts are essential for effectively integrating solar energy sources.

An important aim of our work is to find an practical solution for real world applications. Developing countries in the equatorial zone receive excellent annual insolation, and in many cases are rapidly installing PV systems. In these regions, diffuse radiation often dominates and cloud structures exhibit rapid horizontal and vertical evolution. Most solar forecasting work done to date has been applied to mid-latitude sites: we aim to find a practical forecasting system that takes into account tropical atmospheric conditions and can be operated in real time using available data. Our initial application will be to the island of Singapore, which means that we seek forecasts on a timescale of minutes to hours and a spatial scale of a few kilometres. State-of-the-art irradiance forecasts are evaluated by several

* Corresponding author. Tel.: +65 9159 0888.
E-mail address: A0027350@nus.edu.sg (D. Yang).

0038-092X/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.solener.2012.07.029

Nomenclature

I_{Glo}	global horizontal irradiance (GHI) [W/m^2]
I_{Dir}	DNI (DNI) [W/m^2]
I_{Dif}	diffuse horizontal irradiance (DHI) [W/m^2]
I_{CDir}	clear sky direct normal irradiance (DNI) [W/m^2]
Z	zenith angle [degrees or radians]
CC	opaque cloud cover [dimensionless]
K^*	DNI index [dimensionless]
CSF	fraction of clear sky over the entire local sky [dimensionless]
ARIMA	Auto-Regressive Integrated Moving Average
AIC	Akaike information criterion
MBE	mean bias error
RMSE	root mean squared error
LOESS	local polynomial regression fitting
p	order of the autoregressive process
d	order of nonseasonal differences
q	order of the moving average process

authors e.g. (Lorenz et al., 2009; Perez et al., 2010) and found to be improved by the addition of satellite data and numerical weather prediction (NWP) modelling: we seek a method that does not require these inputs to facilitate cheap, rapid computation.

Among the non-NWP approaches currently used for short term solar irradiance forecasts, two appear suitable for our purposes. The first approach utilizes images from satellites (Perez et al., 2010; Hammer et al., 1999; Perez et al., 2007) or ground-based sky cameras (Chow et al., 2011) to forecast cloud cover based on historical cloud motion. Solar irradiance is then derived from the predicted cloud cover distribution. The second approach forecasts solar irradiance based on historical ground solar irradiance measurements and/or cloud cover indices. We favor the second approach in this paper as the imaging and cloud to solar irradiance mapping in the first approach is complex and requires imaged data, while for the second approach, point data and simple time series analyses can be used.

In the literature, many tools and methods to forecast solar irradiance based on ground solar irradiance measurements have been proposed. These include: time series, autoregressive integrated moving average (ARIMA) analyses (Moreno-Muñoz et al., 2008; Reikard, 2009; Martín et al., 2010), artificial neural networks (ANN) multi-layer perceptron model (Paoli et al., 2010; Mellit et al., 2010; Martín et al., 2010), regression analysis (Reikard, 2009), k-Nearest Neighbors' algorithm (Paoli et al., 2010), Bayesian inference (Paoli et al., 2010) and hybrid model using time series and neural networks (Wu and Chan, 2011; Reikard, 2009). Among these methods, time series analysis requires little data preprocessing. Time series ARIMA analyses, for the timescales of interest and for equal processing power, may have higher accuracy than regression analysis, the hybrid model (Reikard, 2009), k-Nearest Neighbours' algorithm or Bayesian inference methods (Paoli et al., 2010). In particular, at 60 min resolution, Reikard (2009) has shown that an ARIMA model dominates all other time series forecasting methods in four out of six test stations used in that study. Therefore, we use time series ARIMA analysis to forecast solar radiation in this work.

Continuous solar irradiance data acquisition is an important component of time series analysis as we need to trace back several steps in time in order to forecast the solar irradiance at next time step. During night time, solar irradiance is zero, a discontinuity that must be considered when forecasting solar irradiance during the period just after sunrise and before sunset. For periods without solar irradiance measurements, data may be inferred from clear sky models and other meteorological parameters. Weather patterns and their accompanying clouds are the most significant atmospheric phenomena affecting solar irradiance at the earth's surface (Brinsfield et al., 1984), so these must be considered by any forecasting system.

Many previous studies have developed empirical relationships to relate cloud cover conditions and solar irradiance. Refs. Brinsfield et al. (1984), Yaramanoglu et al. (1985), and Kimura and Stephenson (1969) studied the impact of scattered clouds on irradiance on an otherwise cloudless day. In these works, large numbers of meteorological data types, such as various irradiance and solar altitude angle, are used to formulate the relationship between cloud cover and solar irradiance. The computation is complex and missing data jeopardizes the calculation.

Ref. Ehnberg and Bollen (2005) relate solar elevation angle to insolation under different cloud cover conditions. A total of nine regression lines are constructed to form a look-up table according to nine levels of cloud cover on the standard Oktas scale, and the results are tested against stochastically-generated cloud simulations.

Refs. Topcu and Onev (1994), Younes and Muneer (2006) and Luo et al. (2010) relate solar irradiance to cloud cover and zenith angle through regression, while Ref. Ododo et al. (1996) develops a relationship between solar irradiance, cloud cover and relative sunshine duration. Ehnberg and Bollen (2005), Topcu and Onev (1994), Ododo et al. (1996), Younes and Muneer (2006), and Luo et al. (2010) have developed simple models to relate cloud cover and solar irradiance but do not consider forecasting techniques. Hence, in this paper, methods to forecast hourly solar irradiance including cloud cover effects using time series analysis are proposed.

Three forecasting methods are proposed, using different types of meteorological data as input parameters, namely, global horizontal irradiance (GHI), diffuse horizontal irradiance (DHI), direct normal irradiance (DNI) and cloud cover. The first method decomposes GHI, into seasonal, trend and irregular components. Trend and irregular components are used as input for a time series analysis ARIMA model (see below) to forecast next hour GHI. The second method forecasts DNI and DHI separately using decompositions, together with an ARIMA model and then an atmospherical formulation is used to relate the two forecast types of irradiance to forecast GHI. The third method studies the effect of cloud cover. An ARIMA model is used to predict cloud transients. A look-up table of GHI is constructed under different cloud cover conditions and zenith angles. By searching the table using forecast cloud cover and the relevant zenith angle, a GHI forecast is obtained.

The paper is organized as follows. Section 2 introduces various time series models along with applicability of these models when applied to meteorological parameters used in our proposed models. In Section 3, three models to forecast hourly global horizontal irradiance are presented. Section 4 consists of test cases using all three proposed models on data collected from two USA weather stations, namely, Miami and Orlando. A detailed comparison of results of the three proposed models is also shown in this section. Section 5 concludes the findings of this work.

2. Time series analysis

A time series is a sequence of observations taken sequentially in time and is called stationary if it has a constant mean level (Box et al., 1994). Both stationary and non-stationary time series forecasting models have been developed, including auto-regressive (AR), moving average (MA), auto-regressive moving average (ARMA), integrated moving average (IMA), ARIMA and seasonal auto-regressive integrated moving average (SARIMA) (Box et al., 1994). Among these models, AR, MA and ARMA are commonly used to forecast stationary trends; the other models are used to forecast non-stationary trends. For detailed derivation of the equations used in this section, please refer to Box et al. (1994).

We use ARIMA models, as these can deal with both stationary and non-stationary time series, and can be used with integrated and moving average process orders. Our search algorithm, based on the Akaike information criterion, AIC (Ricci, 2005), will automatically search for the optimal model which fits the specific time series. We implement our models using a standard R development package (R Development Core Team, 2011) and in particular the polynomial function (Venables et al., 2009).

2.1. Time series analysis ARIMA model

Following Box et al. (1994), we define a *backshift operator* B of the time series as $Bz_t = z_{t-1}$ where z_t and z_{t-1} are two consecutive time series observations. Then B^j can be defined as $B^j z_t = z_{t-j}$. We now define a process called a *white noise process*: for some random shocks (being events such that the time series value departs from the time series average) drawn from a fixed distribution with mean zero and variance σ_a^2, a sequence $a_t, a_{t-1}, a_{t-2} \ldots$, is called the white noise process. The backshift operator and the white noise process describe the intrinsic features of the time series: adjacent observations are dependent. Box et al. (1994) state that certain non-stationary time series can be obtained by autoregressive integrated moving average models (ARIMA):

$$\phi(B)\nabla^d z_t = \theta(B)a_t \tag{1}$$

where

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$
$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

a_t is an independent shock of time step t of the white noise process. $\phi(B)$ is called the autoregressive operator; $\theta(B)$ is called the moving average operator. The operator ∇ is called the difference operator which is equal to $1 - B$. We write this model as ARIMA (p, d, q) where p, d and q are process orders (Box et al., 1994).

2.2. Input parameters and model selection

Section 1 has introduced the four input parameters of time series analysis used in the three proposed models, namely, global solar irradiance for our first model, DNI and diffuse horizontal irradiance for the second model and cloud cover as input for the third model.

2.2.1. Solar irradiance inputs

One limitation of ARIMA models is that they often overlook the physical behavior of time series objects; in our case sunrise and sunset. A study by Moreno-Muñoz et al. (2008) shows that ARIMA can accurately forecast hourly irradiance levels using data from a few previous hours. However, due to the discontinuity of solar irradiance, forecasts immediately adjacent to sunrise and sunset are problematic. Consider an ARIMA model using hours prior to sunrise measurement as input. If the previous hour's measurement is zero, most likely the forecast value will be near zero according to Eq. (1), which is inaccurate if the forecast hour happens to be the hour of sunrise. Similar arguments apply to the hours after sunset. To solve this problem, we consider the seasonal components of solar irradiance time series, starting with the predictable diurnal cycles.

Fig. 1 shows an example of the diurnal cycle of global horizontal irradiance during 2004 December. By using the *LOESS* method (Cleveland et al., 1990) i.e. local polynomial regression fitting, we can decompose a time series into seasonal, trend and irregular components. The *LOESS* regression is a locally-weighted regression of solar

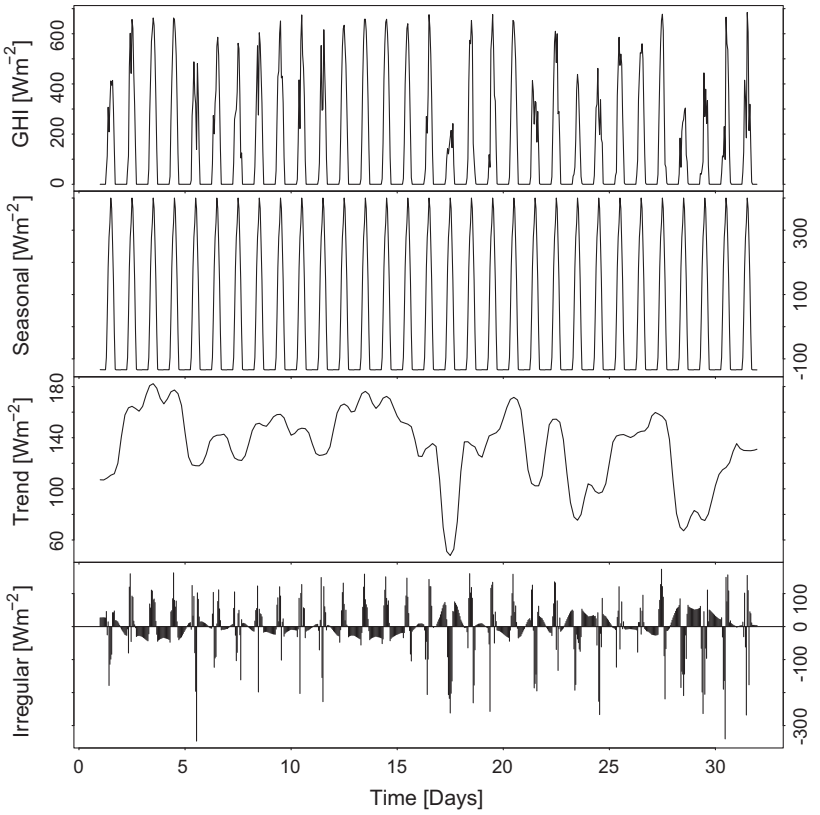

Fig. 1. The decomposition of GHI for Miami 2004 December, with irradiance in Wm^{-2} on the ordinate and day of the month on the abscissa. The top plot is the observed irradiance for 31 days. The second plot shows the diurnal pattern of irradiance. The third and the bottom plots show the trend and irregular components, being the difference between the data and the seasonal component.

irradiance as a polynomial function of time. The seasonal component is found by *LOESS* smoothing the seasonal sub-series (e.g. the series of all solar noon values). The seasonal values are then removed from the solar irradiance time series, the remainders are the trend and irregular components. We then use *LOESS* smoothing to find the trend. Note that for both smoothing processes mentioned above, the smoothing parameters are selected on the basis of knowledge of the time series and on the basis of diagnostic methods (Cleveland et al., 1990). The irregular component is obtained by subtracting the seasonal and trend component from the original series. The decomposition of solar irradiance time series is additive instead of multiplicative. The data we used in this work is hourly, so the seasonal component has a period of 24. Since we apply our ARIMA model month-by-month, the seasonal components adequately capture the sunrise and sunset times even though the sunrise and sunset time vary throughout a year. i.e. in this paper the term "seasonal" refers only to the diurnal cycle, not the summer–autumn–winter–spring cycles.

After the time series is decomposed, the seasonal (diurnal) cycle is subtracted from the original solar irradiance time series, so that the residual is the sum of trend and irregular components. The residual is then used to create a forecast using ARIMA. The parameters of the ARIMA model are estimated based on the Akaike information criterion (AIC). AIC is a measure of the goodness of fit of a statistical model, and is implemented in many standard statistics software packages.

Since the time of sunrise and sunset varies by a few minutes throughout the year, we cannot rely on a single decomposition for an entire year for forecasting purposes: we do one decomposition per month which proves to be adequate. The parameters (process orders p,d,q) of the time series model are not fixed, since we assume, *a priori*, that higher order phenomena drive the irradiance behavior. Therefore, for an optimal solution, a complete time series analysis should be run before each point forecast to determine the process orders (p, d, and q) for optimal ARIMA models.

2.2.2. Cloud cover inputs

Fig. 2 is a typical example of a cloud cover time series, ordinate represents amount of sky dome covered by clouds. It can be seen that cloud cover data does not carry any seasonal pattern. Therefore we use the ARIMA model directly to forecast cloud cover.

Continuous data is used during the model building to preserve the temporal behavior of cloud transients. However, when we develop the regression lines for irradiance using cloud cover and zenith angles (details in Section 3.3), only daytime values of cloud cover data are used, as night time irradiance is known to be zero at all times. Due to the uncertainties in cloud formation and movement, there is no

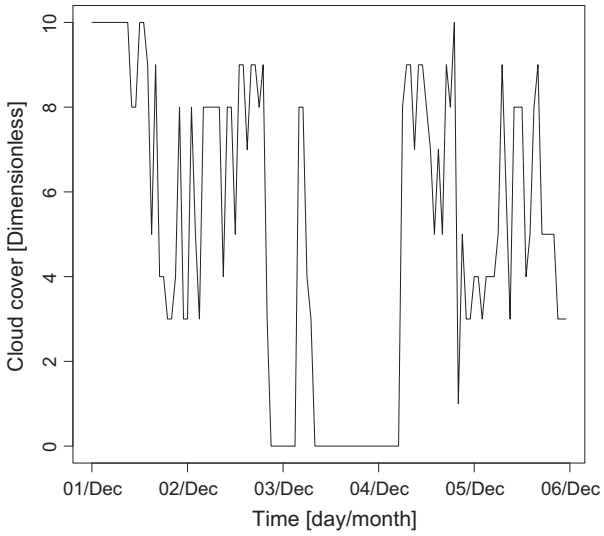

Fig. 2. An example of cloud cover time series plot for Miami (2004 December 1–5). Source: National Solar Radiation Data Base (2001).

predetermined set of time series process orders. As the ARIMA model computation run time is only a few seconds, our ARIMA model building is carried out before each forecast to search the process order parameter space, i.e., to minimize the errors, we iterate the computation for each ARIMA model process order by taking in new observed values before each forecast.

3. Proposed models

In this work the third typical meteorological year (TMY3) data from National Solar Radiation Data Base (2001) is used.[1] The TMY data set provides an annual dataset that holds hourly meteorological values that typify conditions at a specific location over a longer period of time, such as 30 years.

The TMY dataset provides a large variety of hourly meteorological parameters such as solar irradiance (global horizontal irradiance, direct normal irradiance and diffuse horizontal irradiance), cloud cover, position of the sun, temperature, humidity, pressure, wind speed and wind direction. All these parameters are inter-related in a complex physical interaction: our time series analysis aims at a phenomenological approximation of their behavior.

Multiple linear regression analyses were carried out to test for statistical significance of the seven parameters in relation to GHI. We use a backward elimination method (Faraway, 2002) to select the most relevant parameters by examining the *P*-values. The parameter (predictor) with highest *P*-value greater than 0.001 is removed (small *P*-value represent higher relevancy). The elimination process stops when all the parameters are statistically significant. Table 1 shows the *P*-values of selected seven meteorological parameters through out the selection process. The raw data sets used in these regression models are Orlando 2005 October, and Miami 2004 December. Three meteorological parameters show high relevancy to global horizontal irradiance: DNI, diffuse horizontal irradiance and cloud cover.

Global horizontal irradiance is represented by the sum of the vertical component of DNI, DHI and ground-reflected solar irradiance (cf. e.g. Masters, 2004). The ground-reflected irradiance does not concern us because we try to predict the irradiance on a horizontal surface. In a future analysis we aim to consider reflections from clouds however, as these may be significant, especially in tropical regions. We represent global horizontal irradiance I_{Glo} as:

$$I_{\mathrm{Glo}} = \cos Z \times I_{\mathrm{Dir}} + I_{\mathrm{Dif}} \tag{2}$$

where I_{Dir} is DNI, I_{Dif} is diffuse horizontal irradiance and Z is the zenith angle. In this equation, $\cos Z \times I_{\mathrm{Dir}}$ gives the vertical component of DNI.

Cloud cover (CC) correlates with the ratio of DNI and clear sky DNI I_{CDir}. As clear sky DNI is predictable, the fluctuation of ground measured DNI is mostly due to the cloud cover phenomenon. We define two parameters, DNI index K^* and clear sky fraction CSF as:

$$K^* = \frac{I_{\mathrm{Dir}}}{I_{\mathrm{CDir}}} \tag{3}$$

$$\mathrm{CSF} = \frac{10 - \mathrm{CC}}{10} \tag{4}$$

K^* represents the fraction of DNI reaching the ground while CSF gives the amount of clear sky as a fraction of the entire local sky territory. The relationship between these two parameters using an example of available data (National Solar Radiation Data Base, 2001) is shown in Fig. 3. In the plot, CSF takes only discrete values due to discrete cloud cover data on a scale of 0–10. The positive correlation of 0.94 between these two parameters shows that fraction of DNI increases as clear sky territory increases, as expected.

By considering the TMY3 data set, three forecast models are proposed. Fig. 4 shows the flow chart of the three proposed models.

3.1. Model 1

In this model, we use a conventional ARIMA model with the addition of a diurnal irradiance pattern as a periodic parameter to forecast GHI. Measured hourly GHI data is first decomposed into three additive components, namely, seasonal, trend and irregular components. The seasonal component (diurnal irradiance pattern) is then subtracted from the measured GHI. The residual of this subtraction (trend component plus irregular component) is used to forecast for the next hour's residual through

[1] We note that TMY data may contain temporal discontinuities and otherwise not embody the precise characteristics of real data, but we consider it nevertheless adequate for the purposes of algorithm development used here.

Table 1
P-values for multiple linear regression models for Orlando 2005 October, and Miami 2004 December using backward elimination, eliminated parameters are represented by –.

Step	Diffuse irradiance	Direct irradiance	Cloud cover	Dry-bulb	Pressure	Wind speed	Relative humidity
Orlando							
1	$<2\times10^{-16}$	$<2\times10^{-16}$	5.13×10^{-8}	0.083	0.128	0.040	0.311
2	$<2\times10^{-16}$	$<2\times10^{-16}$	7.00×10^{-8}	0.070	0.044	0.052	–
3	$<2\times10^{-16}$	$<2\times10^{-16}$	3.69×10^{-8}	–	0.010	0.109	–
4	$<2\times10^{-16}$	$<2\times10^{-16}$	6.21×10^{-8}	–	0.007	–	–
5	$<2\times10^{-16}$	$<2\times10^{-16}$	5.32×10^{-7}	–	–	–	–
Miami							
1	$<2\times10^{-16}$	$<2\times10^{-16}$	2.69×10^{-6}	0.584	0.712	0.731	0.948
2	$<2\times10^{-16}$	$<2\times10^{-16}$	1.03×10^{-6}	0.583	0.705	0.734	–
3	$<2\times10^{-16}$	$<2\times10^{-16}$	1.04×10^{-6}	0.592	0.754	–	–
4	$<2\times10^{-16}$	$<2\times10^{-16}$	1.05×10^{-6}	0.657	–	–	–
5	$<2\times10^{-16}$	$<2\times10^{-16}$	5.49×10^{-7}	–	–	–	–

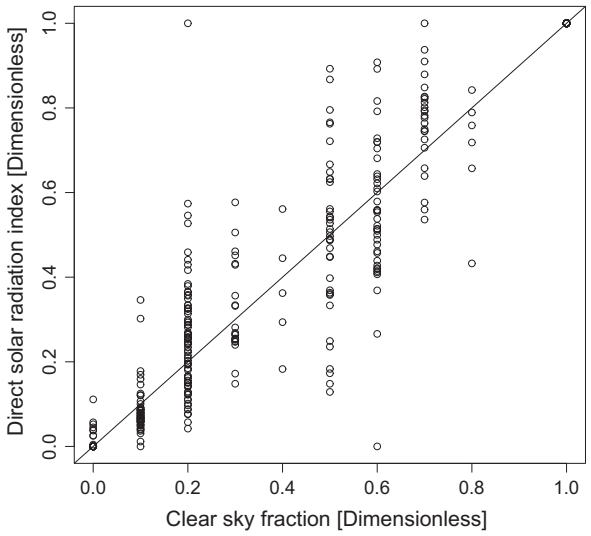

Fig. 3. The relationship between DNI index K^* and clear sky fraction *CSF* for Miami 2004 December. The plot shows a positive correlation of 0.94 between the two parameters.

ARIMA. The final step is to add back the seasonal component to the forecast residual time series to obtain the forecast GHI.

3.2. Model 2

Time series analyses have intrinsic limitations when it comes to data with rapidly-fluctuating uncertainties, especially in short term forecasts such as hourly solar irradiance forecasts. Reikard (2009) has shown that when a particular model (like our model 1), the longer the timescale, the greater the forecast error. A long timescale forecast is associated with more meteorological information being lost than a small timescale forecast. Thus, when only hourly data is available in our study, we seek to improve our forecasts by considering additional input data.

Eq. (2) separates GHI into DNI and DHI, suggests that an improved forecast model would be one that considers DNI and DHI separately. In our second model then, two separate decompositions are carried out using these parameters as time series objects. The residuals are then used as inputs for separate ARIMA analyses. The forecast DNI and DHI are then combined through Eq. (2) to obtain the global irradiance forecast.

3.3. Model 3

Unlike the previous two models, the third proposed model explicitly considers the effect of cloud cover on solar irradiance. To understand cloud cover effects, a physical atmospheric model would ideally be considered. Such atmospheric models are generally difficult to construct and compute in real time. Instead, regression analysis is widely adopted to analyze the relationships among meteorological parameters. Cloud cover data crudely quantifies the amount of sky covered by clouds or other obscuring phenomena at each time step (National Solar Radiation Data Base, 2001). Fig. 5 shows a scatter plot of I_{Glo} versus cosine of zenith angle at each cloud cover condition for our TMY data, i.e., cloud cover quantified on a scale of 0 to 10, where 0 means no clouds. We hypothesise that under different cloud cover condition, GHI follows higher order polynomials of cosine of zenith angle at a specific time. By inspection of the GHI transient with zenith angle, we infer that a quadratic polynomial is not sufficient to describe the cloud-zenith relationship, while polynomial of order four and above may consist of more parameters than can be justified by the data. Therefore we use a polynomial of order three.

The polynomial selected to correlate GHI and the zenith angle is given by Eq. (5):

$$I_{\mathrm{Glo}} = \alpha_0 + \alpha_1 \times \cos Z + \alpha_2 \times \cos^2 Z + \alpha_3 \times \cos^3 Z \qquad (5)$$

where α_0, α_1, α_2 and α_3 are regression parameters. Although Table 2 reports several insignificant regression predictors for this polynomial, we shall use them to preserve regression models' consistency for all cloud cover conditions. Following Eq. (5), a set of 11 regression lines are constructed. These 11 regression lines correspond to 11 cloud

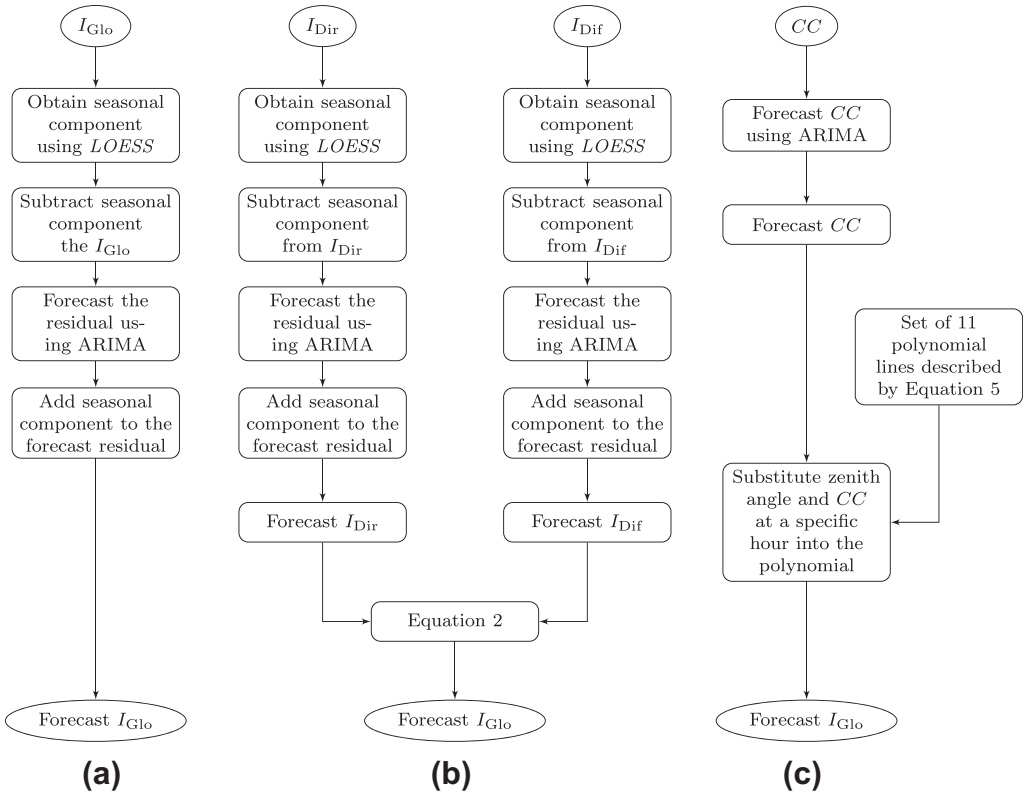

Fig. 4. Flow chart of forecasting methods: (a) use irradiance to forecast next hour solar irradiance through decomposition and ARIMA, (b) forecast DNI and diffuse horizontal irradiance separately using decomposition and ARIMA model and then combine the two forecast irradiance together to predict the irradiance using Eq. (2), and (c) use ARIMA to predict the cloud transients; feed the forecast cloud cover together with zenith angle into nonlinear regression described by Eq. (5) to obtain the solar irradiance forecast.

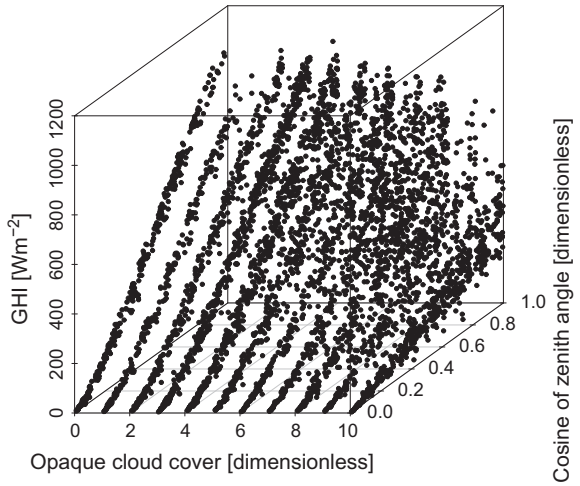

Fig. 5. Scatter plot of I_{Glo} versus cosine of zenith angle at each cloud cover condition.

cover possibilities (on the 0–10 scale), and thus a look-up table is created.

With the help of this look-up table, once the next hour cloud cover is determined using ARIMA model, we obtain the next hour's GHI forecast by substituting the zenith angle at that specific hour, which is a deterministic parameter, into the corresponding regression line.

4. Case studies and discussion

To test the veracity of the three models proposed, two case studies have been made. Data from Miami (25°47′16″N, 80°13′27″W) and Orlando (28°32′37″N, 81°22′22″W) retrieved from (National Solar Radiation Data Base, 2001) is used in this section. All the forecast values are true out-of-sample forecasts, in that they use only data prior to the start of the forecast horizon. The models are estimated over history, the points of the next time step is forecast, and the forecast values are compared with the actual ones. The procedure is iterative until forecasts have been run over the monthly data set. The TMY3 data set consists 12 months from different years, thus, each month is taken as an individual data set for forecasting purposes.

4.1. Forecast using Model 1

Using the techniques introduced earlier, GHI is used as model input to perform the forecast. Forecasting for the entire TMY data set is carried out. As some months in TMY are not continuous but from different years, each month is used independently. We train the model using the first week of a month to obtain the decomposed series. Residuals are then obtained by subtracting the diurnal pattern from the monthly data. After that, ARIMA process

Table 2
P-values of linear regression models of higher order polynomials using TMY3 Miami data set.

Cloud cover	Parameters of polynomial of order 3: $I_{Glo} = \alpha_0 + \alpha_1 \times \cos Z + \alpha_2 \times \cos^2 Z + \alpha_3 \times \cos^3 Z$					
	α_0	α_1	α_2	α_3		
0	0.223	5.85×10^{-15}	$<2 \times 10^{-16}$	$<2 \times 10^{-16}$		
1	0.985	1.66×10^{-4}	$<2 \times 10^{-16}$	$<2 \times 10^{-16}$		
2	0.075	5.11×10^{-10}	3.42×10^{-14}	1.32×10^{-10}		
3	0.079	1.09×10^{-10}	4.34×10^{-9}	1.63×10^{-6}		
4	0.235	9.17×10^{-6}	1.51×10^{-4}	3.94×10^{-3}		
5	0.593	3.40×10^{-3}	1.45×10^{-4}	1.38×10^{-3}		
6	0.657	0.327*	1.02×10^{-4}	1.00×10^{-3}		
7	0.684	0.024	0.028	0.105*		
8	0.881	0.392*	3.4×10^{-3}	3.27×10^{-3}		
9	0.791	0.209*	0.068*	0.038		
10	0.347	3.21×10^{-4}	0.545*	0.399*		
	Parameters of polynomial of order 4: $I_{Glo} = \alpha_0 + \alpha_1 \times \cos Z + \alpha_2 \times \cos^2 Z + \alpha_3 \times \cos^3 Z + \alpha_4 \times \cos^4 Z$					
	α_0	α_1	α_2	α_3	α_4	
0	0.191	1.03×10^{-5}	0.012	0.559*	0.561*	
1	0.334	1.37×10^{-3}	0.061*	0.852*	0.124*	
2	0.990	0.194*	8.91×10^{-5}	0.008	0.057*	
3	0.550	0.04*	9.17×10^{-3}	0.104*	0.284*	
4	0.713	0.165*	0.069*	0.250*	0.417*	
5	0.749	0.203*	0.250*	0.617*	0.904*	
6	0.857	0.227*	0.765*	0.670*	0.403*	
7	0.820	0.754*	0.160*	0.301*	0.394*	
8	0.711	0.301*	0.928*	0.710*	0.471*	
9	0.565	0.668*	0.107*	0.162*	0.246*	
10	0.978	0.454*	0.376*	0.326*	0.275*	
	Parameters of polynomial of order 5: $I_{Glo} = \alpha_0 + \alpha_1 \times \cos Z + \alpha_2 \times \cos^2 Z + \alpha_3 \times \cos^3 Z + \alpha_4 \times \cos^4 Z + \alpha_5 \times \cos^5 Z$					
	α_0	α_1	α_2	α_3	α_4	α_5
0	0.618	0.077*	0.083*	0.334*	0.428*	0.392*

* Denotes parameter is statistically insignificant at a level of 0.05.

orders and parameters are selected and trained using the first week's residual based on AIC. We only perform a single future point forecast, i.e., the forecast is for next hour using the trained parameters. However, the training and forecasting iterates for the entire month. The final step is to add the diurnal pattern to the forecast residual series to obtain the forecast irradiance. The detailed procedure for the simulation is shown in Fig. 6.

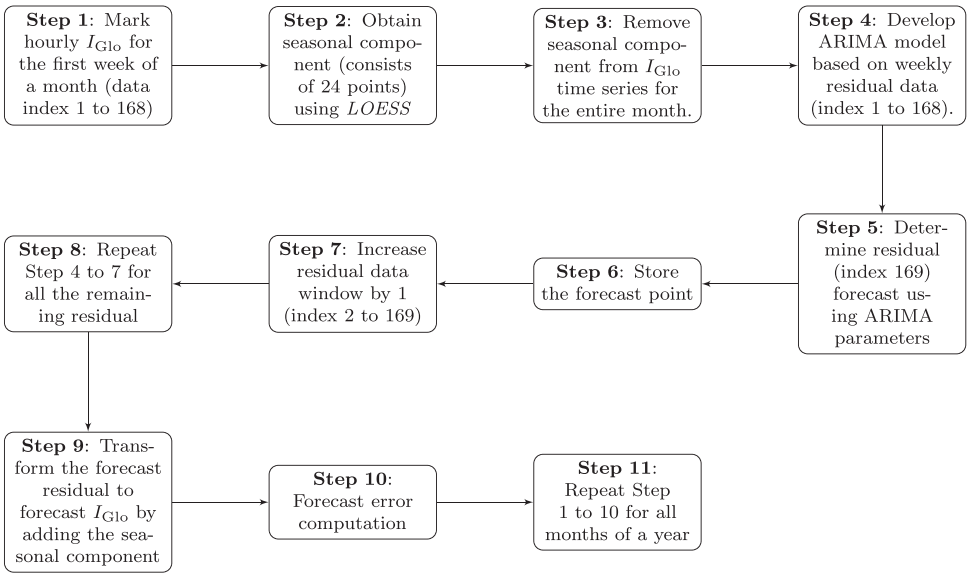

Fig. 6. Simulation of forecasting model 1.

4.2. Forecast using Model 2

The iterative procedures of direct normal irradiance and diffuse horizontal irradiance forecast is identical to the forecast for GHI. However, in this model, DNI and DHI are used individually as model inputs. Then the two types of forecast irradiance are used to construct GHI through Eq. (2).

4.3. Forecast using Model 3

Look-up tables are developed according to method described in Section 3.3. The empirically-determined coefficients using the complete TMY3 data set for Miami is shown in Table 3. The algorithm used in this study can be found in R "polynom" package (Venables et al., 2009). Similarly, the empirical determined coefficients for Orlando is shown in Table 4.

In Fig. 7, GHI is represented as a function of the zenith angle for all 11 possible values of cloud coverage. The topmost regression line shows a situation when cloud cover is zero, which corresponds to clear sky situation; whereas the bottommost regression line shows a situation when cloud cover is 10, which corresponds to completely covered sky condition. Regression lines representing cloud cover 1–9 fall in between the two lines.

Table 3
Correlation coefficients for Miami. *cc* is opaque cloud cover and α_0 to α_3 are the regression coefficients.

cc	α_0	α_1	α_2	α_3
0	−6.963	441.948	1462.358	−868.380
1	0.106	237.797	1882.965	−1215.547
2	−11.997	413.974	1213.413	−667.429
3	−13.465	477.583	1021.191	−544.869
4	−13.692	457.286	898.256	−440.671
5	−7.689	365.900	1066.004	−566.029
6	8.198	154.058*	1355.005	−712.482
7	−7.615	359.399	778.865	−361.263*
8	3.365	166.766*	1279.514	−810.311
9	−6.369	258.566	840.045*	−601.780*
10	−12.134	406.799	−155.162*	139.128*

* Denotes parameter is statistically insignificant at a level of 0.05.

Table 4
Correlation coefficients for Orlando.

cc	α_0	α_1	α_2	α_3
0	−29.365	679.155	876.178	−525.818
1	−13.651	451.225	1291.231	−798.348
2	−19.031	567.432	799.903	−414.411
3	−29.910	683.499	556.550	−286.267*
4	−28.637	581.806	791.531	−485.901
5	−15.907	445.281*	914.465*	−524.176*
6	−11.255	371.442*	572.831*	−151.761*
7	−44.565	798.151	−741.576*	701.471*
8	−24.478	512.575*	316.941*	−139.145*
9	−15.719	451.240*	47.154*	113.185*
10	−5.540	257.107	227.621*	−153.208*

* Denotes parameter is statistically insignificant at a level of 0.05.

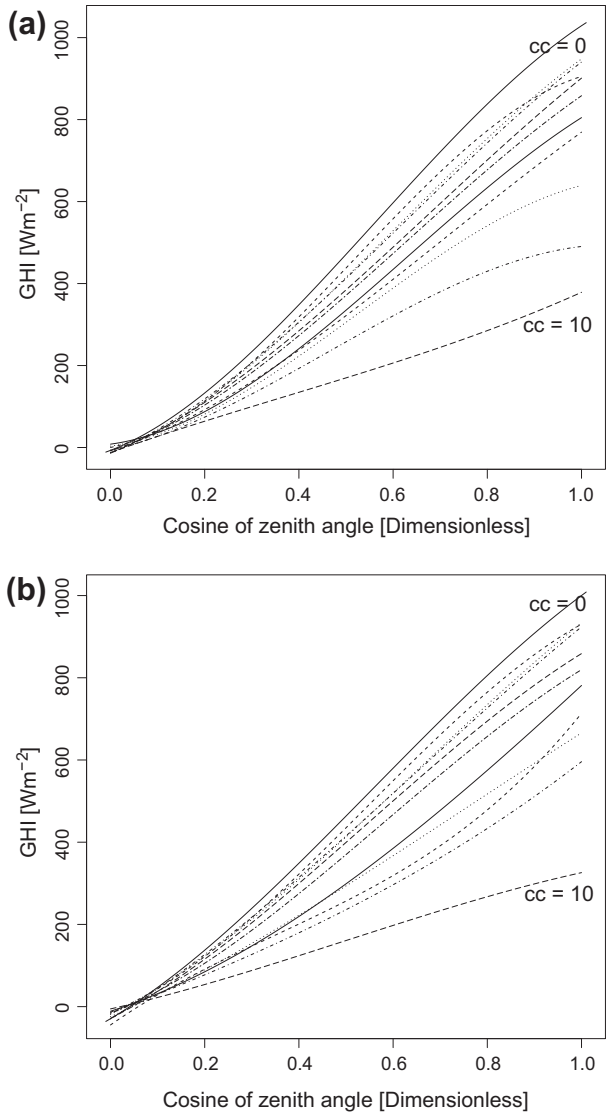

Fig. 7. The relationship between irradiance and the zenith angle at different cloud coverage: (a) Miami, and (b) Orlando, governed by Eq. (5).

After constructing a look-up table, a time series analysis ARIMA model is applied for cloud cover. After obtaining the cloud cover forecast, by searching the table using the forecast cloud cover and the respective zenith angle, we can obtain the GHI forecast.

4.4. Discussion

The criteria used to evaluate the forecast accuracy are the mean bias error (MBE) and root mean squared error (RMSE), defined as:

$$\mathrm{MBE} = \frac{1}{n}\sum_{t=1}^{n}(F_t - A_t) \tag{6}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(F_t - A_t)^2} \tag{7}$$

Table 5
Forecast MBE for hourly TMY3 data at selected sites.

Fcst period	Mean observed I_{Glo}	Model 1	Model 2	Model 3	Persist.
		MBE for Miami			
1995 January 8–31	329.67	−7.27	−8.33	−1.58	−3.39
1981 February 8–28	354.93	−4.31	−6.55	−3.39	−0.59
1988 March 8–31	414.93	−10.79	−13.38	−1.27	−1.77
2001 April 8–30	455.80	15.85	25.03	0.54	−1.27
1990 May 8–31	392.78	−2.55	0.71	1.17	−2.60
1978 June 8–30	382.57	−0.56	−6.56	3.27	−1.14
1990 July 8–31	404.75	0.47	2.59	2.36	−0.84
1994 August 8–31	421.61	−1.09	2.07	1.86	−2.37
1978 September 8–30	387.54	−6.11	−4.67	1.06	−0.67
1994 October 8–31	359.95	−1.81	−5.78	1.06	−2.42
1977 November 8–30	287.14	−1.48	8.22	−0.13	−1.47
2004 December 8–31	296.22	4.43	4.50	−0.62	−4.06
Overall	374.13	−1.28	−0.19	0.39	−1.91
		MBE for Orlando			
1995 January 8–31	281.99	−10.90	−6.50	−10.39	−3.17
1991 February 8–28	328.05	−3.05	−7.68	−1.77	−1.15
2001 March 8–31	380.46	3.69	5.06	1.62	−3.37
1998 April 8–30	485.73	1.24	9.22	−4.15	−2.33
1994 May 8–31	447.92	−0.72	0.22	1.76	−3.77
1996 June 8–30	352.85	10.02	17.57	3.11	−0.21
1991 July 8–31	370.62	4.36	15.24	3.01	−0.28
2003 August 8–31	387.51	2.15	−6.81	3.50	−2.38
1999 September 8–30	361.19	25.34	24.98	1.27	−4.41
2005 October 8–31	344.16	−3.96	−7.40	−0.45	−1.94
2001 November 8–30	340.65	−1.97	−8.04	−0.67	−7.19
1996 December 8–31	263.25	0.10	−1.51	−0.19	−0.94
Overall	362.07	2.16	2.86	−0.27	−2.60

where F_t is the forecast value, A_t is the actual value and n is number of forecast points. Errors are calculated using daylight hours only. Tables 5 and 6 show MBE and RMSE for all three proposed models using Miami and Orlando TMY3 dataset respectively. Overall values recorded in the last row of each table are obtained using sum of product of the number of forecast days in a month and monthly mean values. The total is then divided by total forecast days, which is 281 days in this case.

Persistence is used as a benchmark to our proposed models. A persistence "model" is simply the assumption that the value for the next time step is the same as the present value, i.e., $F_t = A_{t-1}$. The MBE and RMSE of persistence model is shown in the last column of Tables 5 and 6 respectively. Perez et al. (2010) states that RMSE is the key performance factor. We can conclude from RMSE table that all three proposed models have significantly lower RMSE than the persistence model. Among the three models, model 3 has the lowest RMSE, followed by model 2. However, the difference between model 1 and 2 are not significant when compared with model 3.

To understand the limitations of each proposed model, monthly MBE and RMSE are not sufficient. As both error terms represent averaged values, errors from individual days should be considered. Two days are selected to closely examine performance of the three proposed models: Miami 2004 December 13 and 2004 December 25.

4.4.1. Case 1

Fig. 8 shows the Miami GHI time series plot on 2004 December 13. Measurements show strong evidence of a clear sky situation through out the day. As such, all cloud cover values of that day are zero (National Solar Radiation Data Base, 2001). Under these situations, model 3 shows a forecast very close to original measurements. It can be seen from Fig. 8 that model 1 and model 2 do not provide any bell shaped curves. To evaluate the situation, the forecast mechanism of model 1 is first considered.

Model 1 uses 168 (a week of hourly data) points prior to the point of forecast to calculate the seasonal (diurnal) pattern through regression fitting. As such, the fitted seasonal pattern is not necessarily a bell shape. After ARIMA forecasting, this seasonal pattern is added back to the forecast residuals. The addition of the seasonal pattern may result in the unexpected forecast error during clear sky condition. Model 2 takes a similar detrend approach to model 1, thus the above analysis applies equally. To resolve this problem, numerical weather prediction (NWP) shall be considered. NWP provides an accurate medium range forecast for cloud motion. By utilizing NWP forecasts for the overall

Table 6
Forecast RMSE for hourly TMY3 data at selected sites.

Fcst period	Mean observed I_{Glo}	Model 1	Model 2	Model 3	Persist.
		RMSE for Miami			
1995 January 8–31	329.67	82.36	73.70	58.95	135.16
1981 February 8–28	354.93	84.82	78.76	35.72	145.27
1988 March 8–31	414.93	113.87	111.14	36.15	166.12
2001 April 8–30	455.80	135.19	136.37	42.37	190.14
1990 May 8–31	392.78	112.10	99.09	25.82	151.22
1978 June 8–30	382.57	123.96	119.26	36.19	161.16
1990 July 8–31	404.75	92.64	94.79	23.83	152.28
1994 August 8–31	421.61	109.26	105.20	23.78	162.32
1978 September 8–30	387.54	128.00	121.39	23.64	170.37
1994 October 8–31	359.95	102.55	100.34	18.92	160.48
1977 November 8–30	287.14	82.21	82.04	14.34	132.77
2004 December 8–31	296.22	108.42	109.17	17.74	154.13
Overall	374.13	106.35	102.69	29.73	156.81
		RMSE for Orlando			
1995 January 8–31	281.99	89.62	80.74	50.91	125.85
1991 February 8–28	328.05	108.35	100.62	39.42	150.14
2001 March 8–31	380.46	136.11	127.42	58.30	176.79
1998 April 8–30	485.73	106.80	107.85	37.21	162.64
1994 May 8–31	447.92	100.43	96.93	29.32	163.23
1996 June 8–30	352.85	131.36	118.08	33.36	160.36
1991 July 8–31	370.62	111.25	109.06	29.10	159.29
2003 August 8–31	387.51	166.15	163.79	38.29	204.89
1999 September 8–30	361.19	140.96	133.45	27.48	179.23
2005 October 8–31	344.16	125.34	120.34	19.72	170.77
2001 November 8–30	340.65	94.60	88.03	17.93	142.38
1996 December 8–31	263.35	90.41	86.70	12.77	130.51
Overall	362.07	116.85	111.18	32.80	160.61

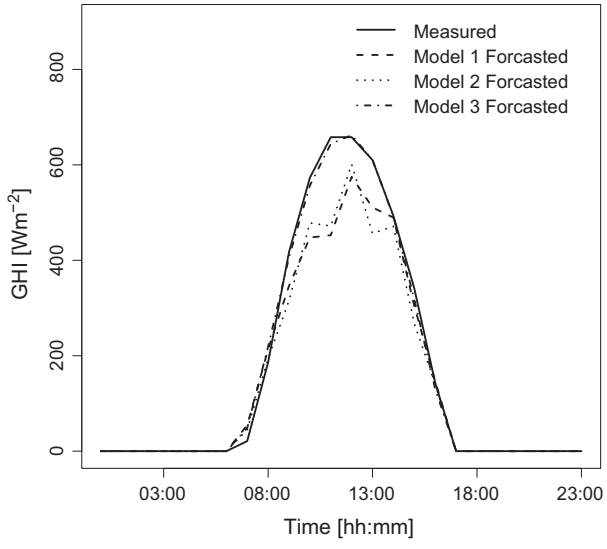

Fig. 8. The irradiance time series plot for Miami 2004 December 13.

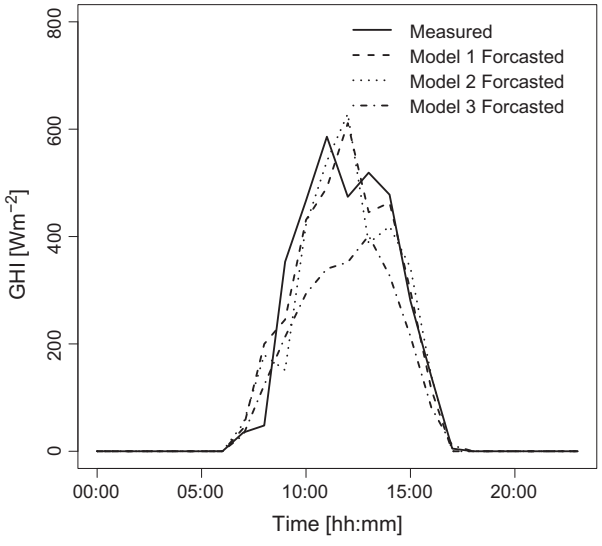

Fig. 9. The irradiance time series plot for Miami 2004 December 25.

situation of the next day, we could select other models such as model 3 to perform hourly forecast for next day.

4.4.2. Case 2

In the case of the data from Miami on 2004 December 25, it is shown that solar irradiance ARIMA forecasts, namely, model 1 and 2, give a better results than the forecast using cloud cover information, model 3, as shown in Fig. 9.

If we look at the time series plot for cloud cover as shown in Fig. 10, the variation of cloud cover is from 7 to 10 for daylight hours on that day (hour 06:00 to 17:00 on the abscissa). On such a cloudy day, the GHI appears

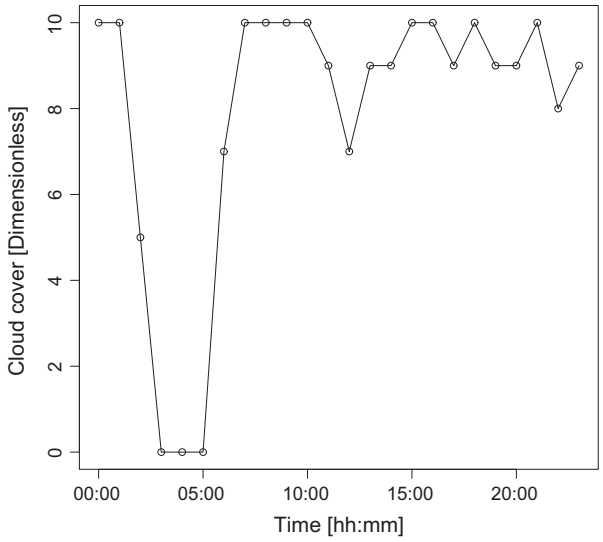

Fig. 10. The time series plot for measured cloud cover condition for Miami 2004 December 25. *Source*: National Solar Radiation Data Base (2001).

to be at similar levels compare to 2004 December 13, Miami, which has zero cloud cover for the whole day. Under this situation, forecasts using cloud cover would give a predicted value which is lower compared to the actual irradiance level.

We suspect that such situation is caused by inaccurate cloud cover data. For partly cloudy skies, an hour can be composed of large or small amounts of sunshine, depending on whether the sun is mostly free of the clouds or occluded by the clouds. Consequently, hourly solar irradiance may depart significantly from actual values for partly cloudy skies conditions (National Solar Radiation Data Base, 2001). As the cloud cover used in the case study is an hourly data collected only at the time indicated (National Solar Radiation Data Base, 2001), namely, at the beginning of each hour, it sometimes cannot represent the hourly average value of cloud cover. Consequently, the forecast solar irradiance may also deviate from actual solar irradiance significantly.

We believe that meteorological parameters, such as cloud cover, can increase the forecast accuracy only if the data set is sufficiently accurate to represent the actual situation. In our case study, we find a need to increase the sampling frequency of cloud cover. Then, cloud cover would better reflect the true value for partly cloudy skies conditions. We suggest that sites with adequate equipment, such as total sky imagers, should collect cloud cover data every five to ten minutes. For sites using experienced human observers, at least two observations need to be made every hour.

5. Conclusions

This paper proposes three models to forecast the next hour global horizontal irradiance. The first two models use solar irradiance as input parameters of the models. Seasonal components of each type of solar irradiance are examined first, and are removed from the original solar irradiance time series. ARIMA models are then used to forecast the residuals. These two models show improved accuracy from a conventional ARIMA forecast. The third model relates global irradiance with a polynomial expression of zenith angle under various cloud cover condition. Solar irradiance forecasts are obtained by feeding the ARIMA forecast cloud cover into a pre-determined look-up table. The merit of the third method is that it requires only low-resolution, ground-based cloud cover data to forecast next hour solar irradiance. The third method is shown to be very useful for its accuracy, especially for situations which lack solar irradiance measurements.

Taken together, these models illustrate practical forecasting methods for use in real time situations with limited real time data. Future work will attempt to further improve forecast accuracy by incorporating higher-resolution imaging, full SARIMA modeling and more input parameters specific to tropical climates.

Acknowledgments

This work is supported by two projects namely "Computational tools for optimal planning and scheduling of distributed renewable energy sources" and "Novel monitoring and control unit for enhanced availability and reliability of solar PV systems Optimization of photovoltaic electricity generation in tropical power grids through irradiance forecasting and system monitoring". Both Projects are funded by the National Research Foundation of Singapore under the Clean Energy Research Programme Grant Nos. NRF2007EWT-CERP01-0954 and NRF2010EWT-CERP001-030.

References

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., 1994. Time Series Analysis: Forecasting and Control. Prentice Hall Inc., Englewood Cliffs, New Jersey.

Brinsfield, R.B., Yaramanoglu, M., Wheaton, F., 1984. Ground level solar radiation prediction model including cloud cover effects. Solar Energy 33, 493–499.

Chow, C.W., 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, 2881–2893.

Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I., 1990. Stl: a seasonal-trend decomposition procedure based on loess. Journal of Official Statistics 6, 3–73.

Ehnberg, J.S.G., Bollen, M.H.J., 2005. Simulation of global solar radiation based on cloud observations. Solar Energy 78, 157–162.

Faraway, J.J., 2002. Practical Regression and Anova using R. <http://cran.r-project.org/doc/contrib/Faraway-PRA.pdf> (accessed 09.11.2011).

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

Kimura, K., Stephenson, D.G., 1969. Solar radiation on cloudy days. In: Transactions American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Ottawa. <http://nparc.cisti-icist.nrc-cnrc.gc.ca/npsi/ctrl?action=rtdoc&an=5211159>.

Lorenz, E., Hurka, J., Heinemann, D., Beyer, H., 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, 2–10.

Luo, L., Hamilton, D., Han, B., 2010. Estimation of total cloud cover from solar radiation observations at lake Rotorua, New Zealand. Solar Energy 84, 501–506.

Martín, 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, 1772–1781.

Masters, G.M., 2004. Renewable and Efficient Electric Power Systems. John Wiley & Sons Inc., Hoboken, New Jersey.

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, 771–782.

Moreno-Muñoz, A., de la Rosa, J.J.G., Posadillo, R., Pallarés, V., 2008. Short term forecasting of solar radiation. In: IEEE International Symposium on Industrial Electronics, Cambridge. pp. 1537–1541.

National Solar Radiation Data Base, Typical Meteorological Year Data Sets. <http://rredc.nrel.gov/solar/old_data/nsrdb/> (accessed 16.02.2011).

Ododo, J.C., Agbakwuru, J.A., Ogbu, F.A., 1996. Correlation of solar radiation with cloud cover and relative sunshine duration. Energy Conversion and Management 37, 1555–1559.

Paoli, C., Voyant, C., Muselli, M., Nivet, M.L., 2010. Forecasting of preprocessed daily solar radiation time series using neural networks. Solar Energy 84, 2146–2160.

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

Perez, R., Moore, K., Wilcox, S., Renné, D., Zelenka, A., 2007. Forecasting solar radiation—preliminary evaluation of an approach based upon the national forecast database. Solar Energy 81, 809–812.

R Development Core Team, 2011. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. ISBN:3-900051-07-0.

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

Ricci, V., 2005. R Functions for Regression Analysis. <http://cran.r-project.org/doc/contrib/Ricci-refcard-regression.pdf> (accessed 29.06.2011).

Topcu, S., Onev, S., 1994. The estimation of hourly total irradiation for cloudy sky in Istanbul. Renewable Energy 4, 223–226.

Venables, B., Hornik, K., Maechler, M., 2009. A Collection of Functions to Implement a Class for Univariate Polynomial Manipulations. <http://cran.r-project.org/web/packages/polynom/polynom.pdf> (accessed 16.04.2011).

Wu, J., Chan, C.K., 2011. Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN. Solar Energy 85, 808–817.

Yaramanoglu, M., Brinsfield, R.B., Muller, R.E., 1985. Estimation of solar radiation using stochastically generated cloud cover data. Energy in Agriculture 4, 227–242.

Younes, S., Muneer, T., 2006. Improvements in solar radiation models based on cloud data. Building Services Engineering Research & Technology 27, 41–54.