Renewable Energy 68 (2014) 421–427

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

ARIMA and regression models for prediction of daily and monthly clearness index

Jamal Hassan*

Khalifa University (KU), P.O. Box 127788, Abu Dhabi, United Arab Emirates

ARTICLE INFO

Article history:
Received 5 July 2013
Accepted 4 February 2014
Available online 12 March 2014

Keywords:
Global and diffuse solar radiation
Regression models
Time series ARIMA models
Clearness index

ABSTRACT

Hourly and daily measurements of total and diffuse solar radiation and sunshine duration are analyzed. Three-year measurements of the available meteorological data in Mosul (latitude 36° 19′ N, longitude 43° 09′ E and 223 m above mean sea level) are used in this study. The present work involves two parts; in the first one, monthly mean daily and hourly of total and diffuse solar radiation are analyzed. The results show that the annual mean of the daily total and diffuse solar radiation is 5.11 and 1.6 kWh/m^2 respectively. 57% of the days of the year are clear, while only 11.5% of the days are cloudy. Several empirical equations for estimating monthly mean daily global and diffuse solar radiation have been developed and compared with other available models. Ratio of average hourly to daily total solar radiation, for each month of the year is studied and compared with the theoretical results. In the second part, time-series model building using Box–Jenkins procedure to the daily clearness index is performed. The ARIMA(2,1,1) is developed for predication of daily clearness index.

1. Introduction

Knowledge of global and diffuse solar radiation at a particular location is required for all solar energy conversion systems. Hourly and daily measurements of the weather parameters such as global, diffuse and sunshine duration are usually the best source of information that provide the starting point for obtaining empirical models for solar radiation estimation. When such measured data is lacking, one option is to estimate solar radiation from empirical equations developed in other locations, preferably sites with similar weather condition to the location under investigation. Iraq, the country located in the middle-east cannot catch up with the increasing demand of electricity not only due to the increase in population but also as a result of the higher demand for electrical power in domestic and industrial sectors. The country possesses a large reservoir of conventional energy resources but has the opportunity, due to its geographical location, to utilize solar energy effectively, promoting a clean environment and developing renewable energy technologies. Several empirical formulas, that goes back more than seventy years ago, have been developed elsewhere to estimate solar radiation using various parameters [1–4]. Iqbal [5] developed several empirical equations which correlate the monthly average daily diffuse and beam radiation with the fraction of maximum possible number of bright sunshine hours for several Canadian cities. Erbs et al. [6] estimated the diffuse radiation fraction for hourly, daily and monthly average global radiation for several sites in the USA. Abdalla et al. [7] studied global solar radiation for the city of Abu-Dhabi, UAE in terms of relative sunshine duration. One year data of measured global and direct solar radiation, is analyzed and compared with the estimated values from NASA's model by Islam et al. [8,9] in the UAE. Al-Riahi et al. [10] developed a clear-day model for the beam transmittance of the atmosphere based on five-year daily global radiation data in Baghdad, Iraq. Several authors successfully applied time-series analysis [11,12] using Box–Jenkins approach for solar radiation data. Hassan et al. [13] obtained an ARMA(2,1) for the global solar radiation in Al-Ain, UAE. They combined a time-series model with a regression model to obtain the optimum predication. Sulaiman et al. [14] and Zaharim et al. [15] used the ARMA Box–Jenkins method to model global solar radiation data. Zeroual [16] used stochastic modeling for daily global solar radiation in Marrakesh, Morocco.

Previously, estimation of solar radiation in many cities in Iraq, including Mosul (the site under investigation in this report) was obtained by using models developed elsewhere that do not often provide accurate outcomes. This study is a first step toward the goal of developing different classical one-parameter-based regression models and time series ARIMA model for the estimation and

* Tel.: +971 2 5018524; fax: +971 2 4472442.
E-mail addresses: jamal.hassan@kustar.ac.ae, dylanjamal@yahoo.com.

http://dx.doi.org/10.1016/j.renene.2014.02.016
0960-1481/

predication of daily and monthly global and diffuse solar radiation in the city of Mosul-Iraq (latitude 36° 19′ N, longitude 43° 09′ E and 223 m above mean sea level). The data used in this paper cover three year period and are recorded at the Meteorological and Solar Energy Station of Mosul University-Iraq. An Eppley pyranometer model 8.48 is used for recording global solar radiation. Similar instrument provided with a shading ring is used for recording diffuse solar radiation. The sensitivity of each device is 1×10^{-3} V/W/m^2. The sunshine duration is measured by Campbell Stokes sunshine recorder with a threshold flux of about 210 W/m^2.

This work presents several empirical models obtained for the estimation of global and diffuse solar radiation from the available hourly and daily data of this location. In addition, a time-series ARIMA model is developed for predication of daily global solar radiation.

2. Theoretical background

2.1. Regression models

Several relations are used to estimate global solar radiation (GSR) and diffuse solar radiation based on the available meteorological parameters. Prescott [17] modified Angstrom's [18] equation for estimation of monthly mean daily GSR ($\overline{H}$) received on a horizontal surface based on monthly mean daily number of hours of possible sunshine $\overline{S}$. The equation is given as:

$$\overline{H}/\overline{H}_o = a + b\left(\overline{S}/\overline{S}_o\right) \quad (1)$$

here $\overline{H}_o$ is the monthly mean daily extraterrestrial solar radiation on a horizontal surface. H/H_o is called clearness index K_t and S/S_o is denoted fraction of sunshine duration F_s. S_o is the monthly mean daily maximum day length of the location while a and b are regression coefficients. The daily values of H_o, on a horizontal surface, and S_o can be calculated using the equations in Ref. [19].

In general, $\overline{H}$, $\overline{H}_d$ and $\overline{S}$ are related through the following relations:

$$\overline{H}/\overline{H}_o = f\left(\overline{S}/\overline{S}_o\right),\ \overline{H}_d/\overline{H}_o = f\left(\overline{S}/\overline{S}_o\right) \text{ and } \overline{H}_d/\overline{H}_o = f\left(\overline{H}/\overline{H}_o\right)$$

which can be written as:

$$\overline{K}_t = f\left(\overline{F}_s\right),\ \overline{K}_d = f\left(\overline{F}_s\right) \text{ and } \overline{K}_d = f\left(\overline{K}_t\right).$$

The form of the function (e.g. linear, quadratic, etc.) in these relations is dictated by the data. There is a relatively simple equation used by Liu and Jordan [20] for estimating hourly mean daily GSR ($\overline{I}$) from monthly mean daily GSR ($\overline{H}$) using the following relation:

$$\overline{I}/\overline{H} = \overline{r}_t = (\pi/24)[\cos w_i - \cos w_s]/[\sin w_s - w_s \cos w_s] \quad (2)$$

where w_s is the hour angle calculated from the equation in Ref. [19] and w_i is the average of hour angle for hour-pairs around noon calculated from the following equation $w_i = (\pi/12)(12 - t_i)$ where t_i is hours from noon time.

All calculations and data fitting, for the regression modeling part, are performed using OriginLab software [21]. The validation of the models and their accuracies were tested by calculating different statistical parameters. These are, mean percentage error (MPE), mean bias error (MBE), root mean square error (RMSE) and the Nash–Sutcliffe equation (NSE), which are described according to the following equations [22]:

$$\text{MPE} = \frac{1}{n}\sum_{1}^{n}\left[\left(\overline{H}_e - \overline{H}_i\right)/\overline{H}_i\right] \times 100 \quad (3)$$

$$\text{MBE} = \frac{1}{n}\sum_{1}^{n}\left(\overline{H}_e - \overline{H}_i\right) \quad (4)$$

$$\text{RMSE} = \sqrt{\left[\frac{1}{n}\sum_{1}^{n}\left(\overline{H}_e - \overline{H}_i\right)^2\right]} \quad (5)$$

$$\text{NSE} = 1 - \left[\sum_{1}^{n}\left(\overline{H}_i - \overline{H}_e\right)^2 / \sum_{1}^{n}\left(\overline{H}_i - \overline{H}_{ave}\right)^2\right] \quad (6)$$

here $\overline{H}_i$ represents $\overline{H}$ or $\overline{H}_d$ and n is the number of observations, $\overline{H}_e$ is the monthly mean daily estimated values of $\overline{H}$ or $\overline{H}_d$ using the obtained models and $\overline{H}_{ave}$ is the average value of $\overline{H}$ or $\overline{H}_d$.

2.2. Time series analysis (ARIMA modeling)

Clearness index (K_t) is a stochastic process which can be modeled using time-series tools such as Box–Jenkins approach [23]. Auto Regressive Moving Average (ARMA) model for any process, such as K_t with orders p,q denoted as ARMA(p,q), are the combination of past values of K_t and past errors. The general equation of ARMA(p,q) for a time series parameter such as clearness index K_t can be written as:

$$K_t = \phi_1 K_t + \phi_2 K_{t-1} + \ldots + \phi_p K_{t-p} + E_t - \theta_1 E_{t-1} - \theta_2 E_{t-2} \ldots - \theta_q E_{t-q}$$

where $\phi_1, \phi_2, \ldots \phi_p$ and $\theta_1, \theta_2 \ldots \theta_q$ are unknown autoregressive and moving average coefficients, respectively. These will be estimated from sample data (K_t in our case). $E_t, E_2, \ldots E_{t-q}$ are statistically independent random shocks that are assumed to be randomly selected from a normal distribution with zero mean and constant variance. Most of the modeling methods including Box–Jenkins approach [22] are applicable on stationary time series. Therefore, to obtain a model for a particular time-series the stationarity of the data need to be checked. A stationary time-series has a quasi-normal distribution with zero mean and constant variance. To check for this, Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) plots are examined [24]. If the ACF of the time series value either cuts off or dies down fairly quickly, then the time series values should be considered stationary. On the other hand, if the ACF dies down extremely slowly, then the time-series values may be considered non-stationary. The autocorrelation at lag k, ACF(k), is the (linear) Pearson correlation between observations k time periods (lags) apart. If the ACF(k) differs significantly from zero, the serial dependence among the observations must be included in the final model. Like the ACF(k), the partial autocorrelation at lag k, or PACF(k), measures the correlation among observations k apart. However, the PACF(k) removes, or "partials out," all intervening lags.

Several methods are used to convert a non-stationary time-series into a stationary one. Differencing the data is one of the efficient ways and usually the first differencing ($\nabla K_t = K_t - K_{t-1}$ in our case) is sufficient for this conversion. In this stage, the obtained model is denoted ARIMA(p,d,q) where "d" refers to the number of differencing. For the obtained differencing time-series the ACF and PACF will be examined again to get some idea of the order (p,q) of the ARIMA model. ACF and PACF are used to determine p orders and q orders of the ARIMA model. For example, for autoregressive