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:

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

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:

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

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–Sutcliffie equation (NSE), which are described according to the following equations [22]:

$$\text{MPE} = \frac{1}{n}\sum_1^n [(\overline{H}_e - \overline{H}_i)/\overline{H}_i] \times 100 \tag{3}$$

$$\text{MBE} = \frac{1}{n}\sum_1^n (\overline{H}_e - \overline{H}_i) \tag{4}$$

$$\text{RMSE} = \sqrt{\left[\frac{1}{n}\sum_1^n (\overline{H}_e - \overline{H}_i)^2\right]} \tag{5}$$

$$\text{NSE} = 1 - \left[\sum_1^n (\overline{H}_i - \overline{H}_e)^2 / \sum_1^n (\overline{H}_i - \overline{H}_{ave})^2\right] \tag{6}$$

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 lags 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

processes ARIMA(p,0,0), the ACF declines exponentially and the PACF spikes on the first p lag. By contrast, for moving average processes ARIMA(0,0,q), the ACF spikes on the first q lags and the PACF declines exponentially.

Usually, several ARIMA models are obtained for the same time-series, then the optimum one can be chosen to represent the data based on some statistical analysis. We have used Minitab-14 and JMulTi version 4 software [25,26] for obtaining the ARIMA models. If the residual term shows a normal distribution behavior with constant variance and zero mean, then it resembles white noise error and there is no need for further ARIMA modeling. If the model is adequate, ACF and PACF plots of the residuals should show all spikes within the 95% Confidence Interval (CI) bounds ($\pm 1.96/\sqrt{N}$) where N is the sample size. In this study we used the following statistical tests to choose the best ARIMA model: i) non-significance of ACFs of residuals using Q test based on χ^2 square statistics. The Ljung–Box Q-statistics can be used to check if the residuals from the obtained ARIMA models behave as a white-noise process [24,27]. This test based on the following equation [27]:

$Q = (n+2)(n)\sum_{k=1}^{k} r_k^2/(n-k)$, where $n = N - d$ ($d = 1$ in this study) k, is the maximum number of lags considered and r_k is the sample ACF at lag k. The value of Q (for k value) will be compared to χ^2 distributed with k-degree of freedom (DF). If $Q < \chi^2_{\mathrm{DF},\alpha}$ where α is the chosen significance level, the values of ACFs will be considered to be equal to zero according to the null-hypothesis [24]. This implies that the chosen ARIMA model gives uncorrelated white-noise residuals that have no correlation with each other. On the other hand, if $Q > \chi^2_{\mathrm{DF},\alpha}$ then the null hypothesis can be rejected and the ARIMA model does not consider being the adequate one. ii) the normality test of the residuals of the selected ARIMA models. The residual error should be similar to white-noise with zero mean and constant variance. iii) checking ACF plot of the residuals; If the model was the adequate one, its ACF values should lies within the 95% of the CI bounds. iv) checking the residual's mean square error (MSE) and sum of square error (SSE).

3. Results and discussion

3.1. Regression models

Monthly mean daily values of several parameters ($\overline{H}_o$, $\overline{H}_d$ and $\overline{H}$) are presented in Fig. 1a. As seen, the variation of these parameters, follow the same pattern. They begin with minimum values in December and increase to maximum values in June, then decrease again from June to December. $\overline{H}$ varies between 2.13 kWh/m^2 in December to 7.99 kWh/m^2 in June, while $\overline{H}_d$ ranges between 0.93 kWh/m^2 in December to 2.39 kWh/m^2 in June. In addition, Fig. 1b shows how the clearness index $\overline{K}_t$, diffuse fraction of global solar radiation ($\overline{K}_d$) and relative sunshine duration $\overline{F}_s$ are related to each other. It is evident that $\overline{K}_t$ is inversely proportional to $\overline{K}_d$ and directly proportional with $\overline{F}_s$ which is considered as an important parameter for estimating total solar radiation. Histogram of the frequencies of daily clearness index (K_t) and diffuse fraction of total solar radiation (K_d) is shown in Fig. 2. The percentage frequency of cloudy days $K_t < 0.34$ [28] is quite low, namely 11.5% whereas that of clear days $K_t > 0.64$ is quite high (57%). The intermediate range of K_t ($0.35 < K_t < 0.64$) represents partly cloudy skies which contributes to 31.1% of the total days of the year. It appears that the sky over Mosul is mostly clear during most days of the year. In addition, the result shows that the upper-cutoff values of K_t is 0.84 which represents a very clear day in this location and for K_d it

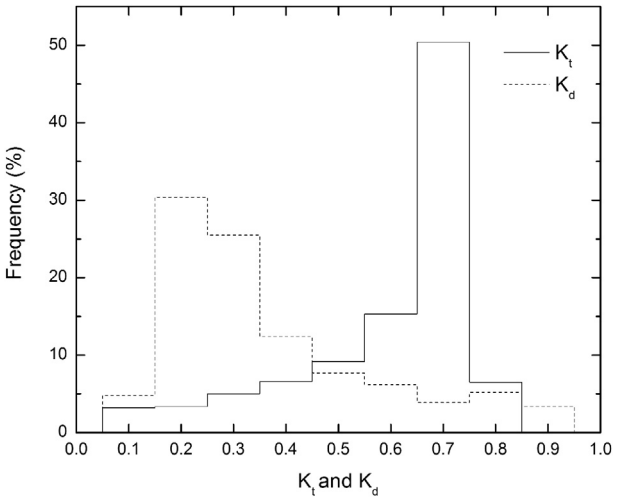

Fig. 2. Histogram of percentage frequency distribution of K_t (solid line) and K_d (dotted line).

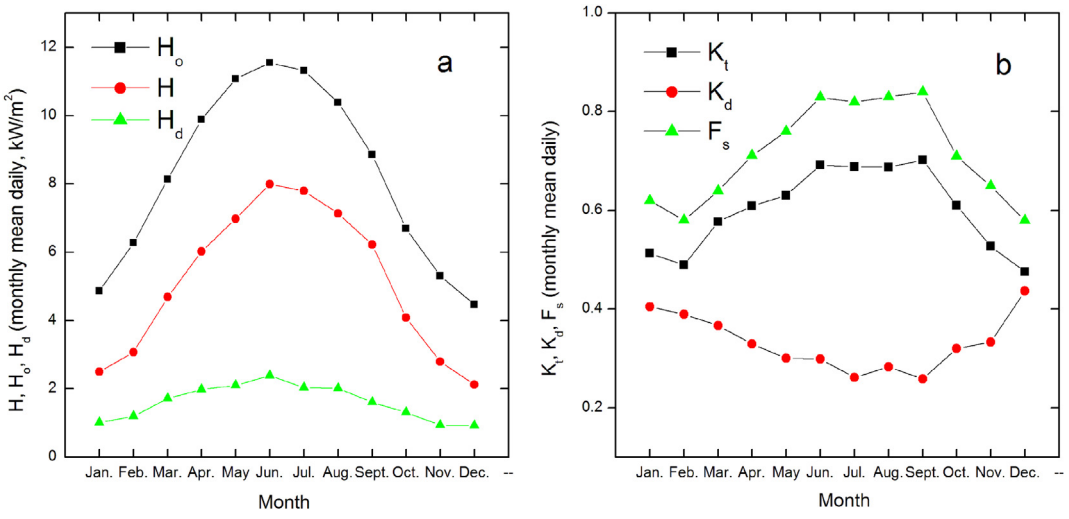

Fig. 1. (a) Monthly mean daily values of total (H), diffuse (H_d) and extraterrestrial (H_o) solar radiation. (b) Monthly mean daily values of clearness index (K_t), diffuse fraction of total solar radiation (K_d) and relative sunshine duration (F_s) for Mosul.

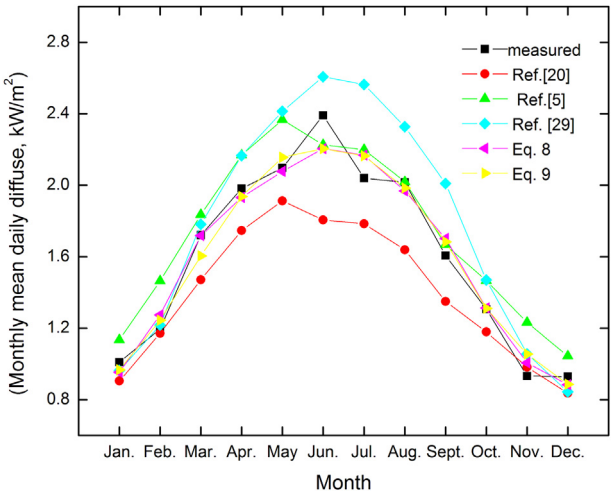

Fig. 3. Measured and estimated diffuse solar radiation using different models as explained within the text.

is equal to 0.94 which is close to a completely overcast day where $K_d = 1$. The lower-cutoff value of both K_t and K_d is equal to 0.05.

Regression equations between pairs of these parameters ($\overline{K}_t - \overline{F}_s$, $\overline{K}_d - \overline{F}_s$ and $\overline{K}_d - \overline{K}_t$) are obtained, from the data, and are summarized below a long with their correlation coefficients.

$$\overline{K}_t = 0.01 + 0.81\overline{F}_s \quad \left(R^2 = 0.97\right) \tag{7}$$

$$\overline{K}_d = 0.71 - 0.53\overline{F}_s \quad \left(R^2 = 0.87\right) \tag{8}$$

$$\overline{K}_d = 0.71 - 0.63\overline{K}_t \quad \left(R^2 = 0.88\right) \tag{9}$$

To study the validity of the obtained empirical equations, the estimated values of $\overline{H}_d$ are obtained using the above relationships. These are compared with their corresponding values of the measured data. The comparison is based on the statistical parameters described in Eqs. (3)–(6). Because there were no empirical models for this location, other equations developed elsewhere [5,20,29] were used for estimating $\overline{H}_d$. Therefore, the developed models are compared with the equations in Refs. [5,20,29] by calculating the statistical parameters using Eqs. (3)–(6). The results are presented in Fig. 3 and Table 1. According to the statistical test results (Table 1), the estimated values are in favorable agreement with the corresponding measured values for the regression model represented by Eq. (8). This is due to the fact this model has the lowest values of MPE, RMSE, MBE and its NSE value is the closest to 1 (the ideal case) compared to the other empirical models. Fig. 3 shows the good agreement between the measured and the estimated values. On the other hand, a close look at the statistical results (Table 1) and Fig. 3 shows that Eq. (9) also gives relatively good agreement with the measured values and it is the second best model for $\overline{H}_d$ estimation. Similar statistical analysis is performed for $\overline{H}$ estimation using Eq. (7) along with the obtained ARIMA model. The results will be discussed in the next section.

Table 1
Statistical analysis for the models used in the estimation of diffuse solar radiation.

Test	Eq. (8)	Eq. (9)	Ref. [5]	Ref. [20]	Ref. [29]
MPE	0.1876	0.3675	10.4577	−11.1347	9.7486
RMSE	0.0911	0.0903	0.1812	0.2594	0.2532
MBE	−0.0021	−0.0008	0.1332	−0.2036	0.1827
NSE	0.9656	0.9662	0.8639	0.7211	0.7341

Monthly mean hourly total ($\bar{I}$) and diffuse ($\bar{I}_d$) solar radiation are calculated (from the measured values) and are plotted versus local time in order to show the diurnal variation of the both parameters in each month of the year as shown in Fig. 4. Both $\bar{I}$ and $\bar{I}_d$ increase with the time of the day and reach their maximum values between 12:00–14:00 h local time and then decrease to their minimum values near the sunset hours. Noon (or near noon) maximum of $\bar{I}$ is due to the minimal ray path in the atmosphere at that time and consequently, minimal attenuation of radiation. Maximum values of ($\bar{I}$) ranges from 368 Wh/m^2 (at 12:00–13:00 h) in December to 958 Wh/m^2 (at 13:00–14:00 h) in June, while for ($\bar{I}_d$) it is between 134 Wh/m^2 (at 12:00–13:00 h) in December and 269 Wh/m^2 (at 13:00–14:00 h) in June. In order to examine whether the data presented here fits Liu-Jordan's theoretical equation (Eq. (2)) or not, the mean measured values at the pair-hours around noon, of the ratio $\bar{r}_t (= \bar{I}/\overline{H})$ are calculated, from the data, for ½, 1½, …5½ hours around noon. The comparison between the measured ratios $\bar{I}/\overline{H}$ and their corresponding theoretical values, calculated from Eq. (2), are shown in Fig. 5. In general, a good agreement can be seen between the measured and theoretical curves for most of the hours around noon time. One also observes that during the summer months (larger values of w_s), a relatively lower values of $\bar{r}_t$ are received around noon (½, 1½, 2½) than in the winter months (smaller values of w_s). On the other hand, higher values of $\bar{r}_t$ are received near sunrise and sunset (around 4½ and 5½). Generally, Liu–Jordan's equation is applicable in estimation of $\bar{r}_t$ values in Mosul.

3.2. Box–Jenkins ARIMA modeling

ACF and PACF values for the mean daily values of the clearness index K_t used in this study are obtained. The ACFs for several lags (k) are outside the 95% confidence Interval (CI) bounds($\pm 1.96/\sqrt{N}$) and they decrease slowly with lag-number. The behavior of the correlations show that the data is not stationary which needs to be converted to a stationary time-series in order to obtain a Box–Jenkins ARIMA model that can fit the data. We have taken the first difference of the data and recalculated the ACF and PACF of the new time-series and presented them in Fig. 6a and b. Note that, the ACF values are within the 95% of the CI bounds except for the first lag. In general, ACF decays in an oscillation form after the first lag. This is an indication that the obtained data (after the first differencing) is stationary. Besides, the PACF plot cuts off just after few lags. This indicates that AR(2) or a higher rank might be an adequate model for the AR part of the final ARIMA model. The second differencing of the time-series led to deterioration of both ACF and PACF behavior, and thus no higher differencing of the data is considered. Therefore ARIMA(p,1,q) will be considered in this study.

We have worked on several ARIMA(p,1,q) models for the data with different parameters (ϕ_i, $i = 1, 2, \ldots p$ and θ_j, $j = 1, 2, \ldots q$). The parameters estimation is based on the least square method with known values of clearness index K_t. Different models are obtained in order to select the best one that describes the daily data of K_t. The selected parameters of the ARIMA(p,1,q) models should have t-values higher than 2.0 in order to be judged significantly different from zero at the 5% level. Furthermore, the coefficients should not be strongly correlated with each other in order to obtain an optimum model for the data. Model coefficients with $P > 0.05$ are insignificant and should be eliminated to avoid obtaining wrong prediction of the data [24].

The AR and MA coefficients of three of models (ARIMA(2,1,1), ARIMA(1,1,1) and ARIMA(1,1,3)) have their P-value smaller than 0.05 (for the chosen $\alpha = 5\%$). This implies that the coefficients of

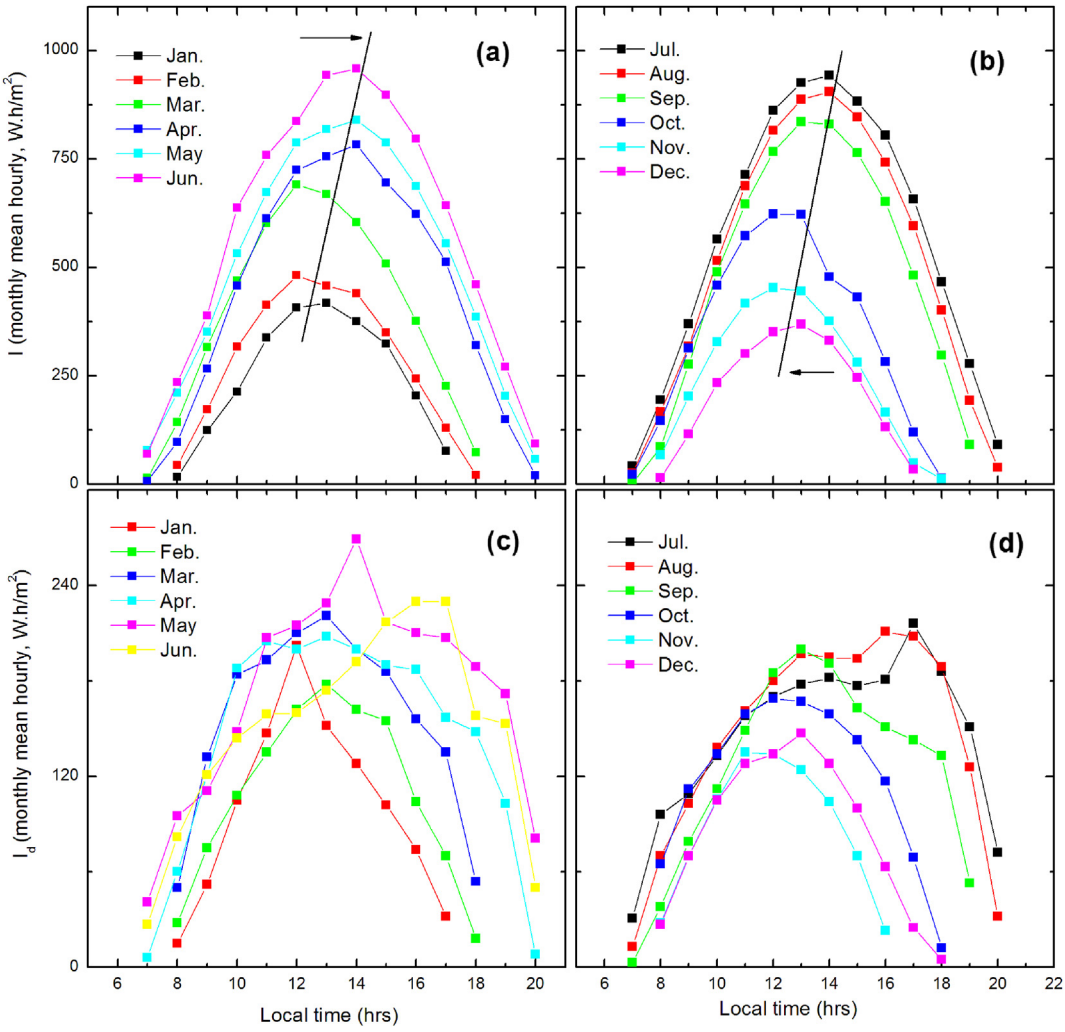

Fig. 4. (a &b) Monthly mean hourly total (I) and (c &d) diffuse (I_d) solar radiation versus local time for every month of the year in Mosul.

these three models are significant according to the null hypothesis which can be rejected if the P-value is larger than 0.05 (for $\alpha = 5\%$). The latter is applicable to ARIMA(4,1,1) as two of its AR parameters, AR(3) and AR(4), have their P-value greater than 0.05. Besides, the t-values for all the parameters of the first three ARIMA models are larger than 2, confirming they are significantly different from zero at 5% level. All the obtained models are used to calculate the daily predicted values of K_t. The results showed that ARIMA(1,1,1) and ARIMA(1,1,3) under-fitted the K_t values while ARIMA(2,1,1) model exhibits a good agreement with the data. ARIMA(2,1,1) residual parameters (MSE = 0.01744, SSE = 6.27873) are the smallest compared with the other models. Therefore, we chose this model for further analysis. The ACF and PACF plots of the residual error for the ARIMA(2,1,1) model are displayed in Fig. 7a and b, respectively. Most, if not all, of the correlations are within the 95% CI bounds. This indicates that the residuals for this model are unrelated and the selected model is suitable for the data.

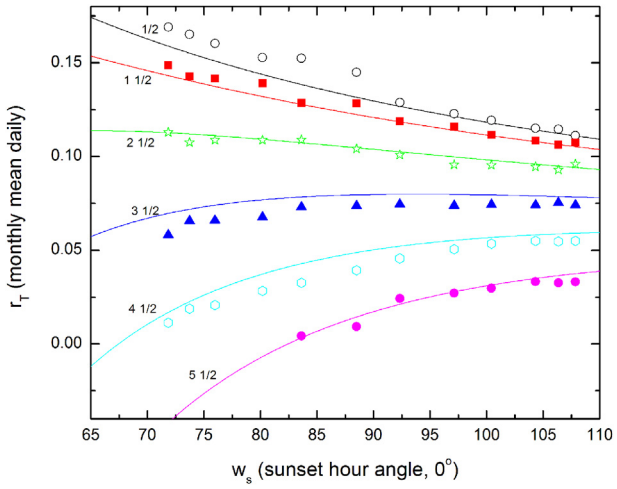

Fig. 5. Measured and theoretical lines, calculated from Eq. (2), of hourly total to daily total solar radiation $r_t = I/H$, at ½, 1½, …..5½ hours around local noon at Mosul.

Ljung-box test results for the ARIMA(2,1,1) residuals are presented in Table 2. As seen from the results, all the P-values (for the selected lag-numbers) are greater than 0.05 and all the *Q_statistics* are smaller than *Chi-squares* ($Q < \chi^2_{DF,\alpha}$) thus indicating that the residual resembles random white noise with low or no correlation with each other. The obtained final model, ARIMA(2,1,1) is given as:

$$\nabla K_t = \phi_1 \nabla K_t + \phi_2 \nabla K_{t-1} + E_t - \theta_1 E_{t-1},$$

since

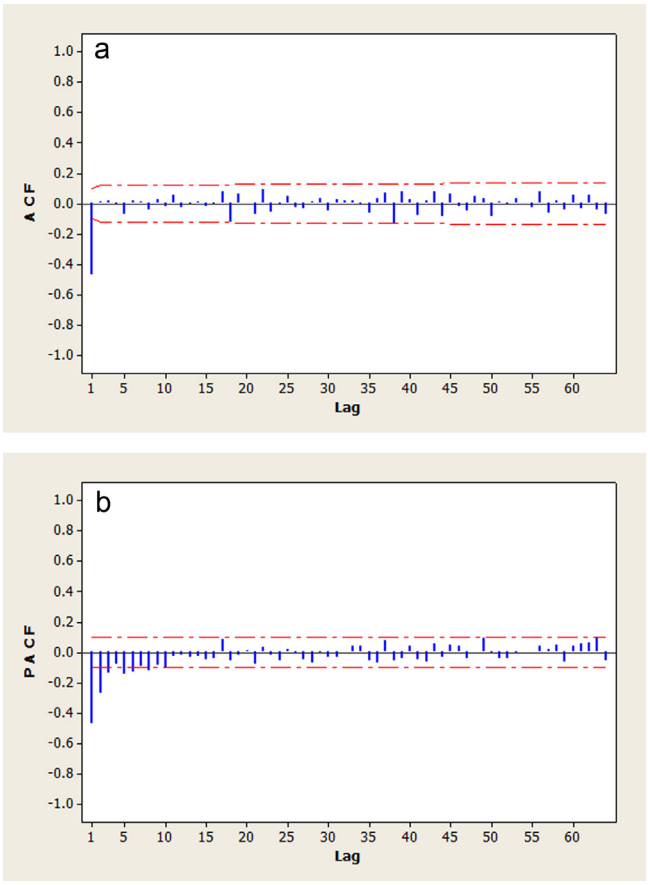

Fig. 6. a) Autocorrelation Function (ACF) of the first differenced-clearness index ∇K_t time series. (b) Partial Autocorrelation Function (PACF) of the first differenced-clearness index ∇K_t time series.

$$\nabla K_t = K_t - K_{t-1},$$

then

$$K_t = K_{t-1} + \phi_1(K_t - K_{t-1}) + \phi_2(K_{t-1} - K_{t-2}) + E_t - \theta_1 E_{t-1}$$

The ARIMA(2,1,1) model establishes that the daily clearness index value (K_t) depends on the value established for the day before (K_{t-1}), with a certain reminder ($\phi_1 = 0.1833$) of the difference between values K_t and K_{t-1}, plus another reminder ($\phi_2 = 0.1230$) of the difference between values K_{t-1} and K_{t-2} plus a normal random variable (E_t) and its reminder ($\theta_1 = 0.9286$) of the moment before. The random variable (E_t) gives a specific mean ($m = 0.0009$) and standard deviation ($\sigma = 0.1315$).

To further check the validity of the obtained ARIMA(2,1,1) model for the daily values, the model is used for prediction of daily values of K_t and the result is incorporated in Fig. 8. It shows a relatively good agreement between the model and the data except for the extreme values of K_t (very high or very low K_t values). It is important in this point to emphasize that this is a typical limitation of ARIMA models. They are designed originally to deal with data of normal distribution while often solar radiation data do not have such distribution. This limitation can be solved by using different techniques [30–32] such as Gaussian Mapping or Markov Transition Matrices (MTM) which will not be investigated in this manuscript. These methods often require complex modeling and data analysis.

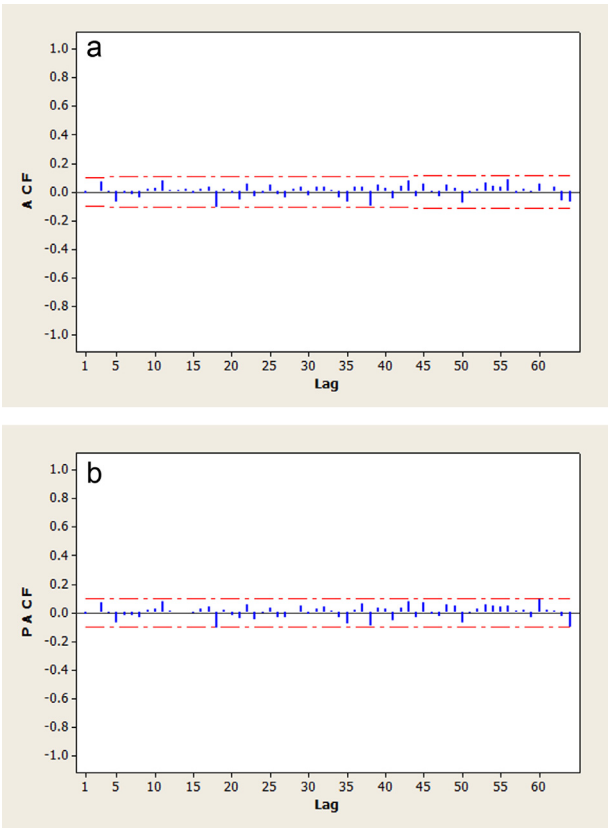

Fig. 7. a) Autocorrelation Function (ACF) of the ARIMA(2,1,1) residuals. (b) Partial Autocorrelation Function (PACF) of the ARIMA(2,1,1) residuals.

Table 2
Ljung–Box test for ARIMA(2,1,1) model for $\alpha = 0.05$ and DF = $k - (p + d + q)$.

Lags (k)	12	24	36	48
DF	8	20	32	44
Chi-square (DF, α)	15.51	31.41	46.19	60.48
Q Statistic	6.9	15.2	21.6	35.3
P-Value	0.553	0.765	0.918	0.823

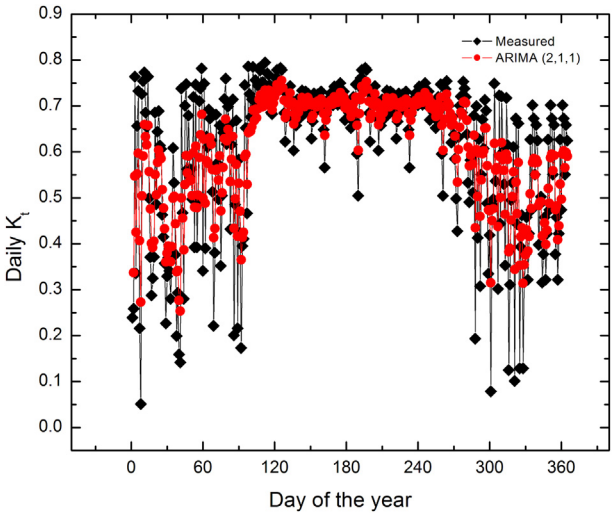

Fig. 8. Measured daily values of K_t with ARIMA(2,1,1) prediction.

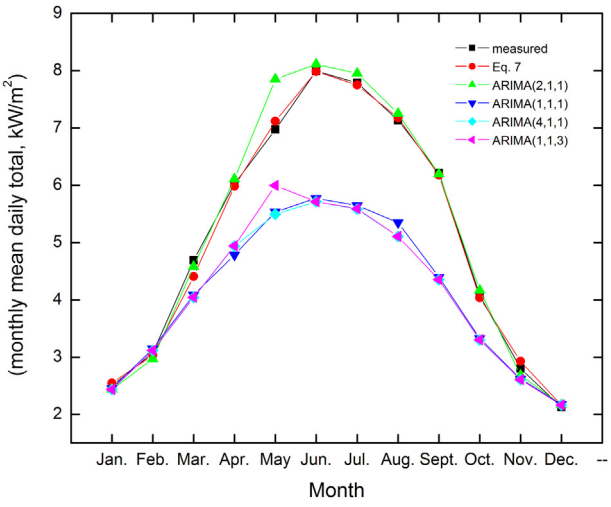

Fig. 9. Measured and estimated total solar radiation using different models used in this study.

Table 3
Statistical analysis for the models used in the estimation of total solar radiation.

Test	Eq. (7)	ARIMA(211)	ARIMA(111)	ARIMA(411)	ARIMA(113)
MPE	−1.8901	0.9045	−15.4165	−16.1644	−16.1136
RMSE	0.1574	0.2714	1.3202	1.3579	1.3554
MBE	−0.1064	0.0908	−1.0087	−1.0466	−1.0440
NSE	0.9942	0.9829	0.5945	0.5710	0.5726

To investigate the possibility of using ARIMA model in estimating monthly values, each model is used in generating daily clearness indices for a year then monthly mean daily total solar radiations $\overline{H}$ are obtained from these values. Finally, these are compared with the corresponding values estimated from the regression model (Eq. (7)) and the result is presented in Fig. 9.

Statistical analysis based on Eqs. (3)–(6) are performed for these models and presented in Table 3. It is obvious that Eq. (7) has the lowest values of MPE, RMSE, MBE and its NSE value is the closest to 1 (the ideal case) compared with the obtained ARIMA models. Although, the ARIMA(2,1,1) model is the second best model to estimate the $\overline{H}$, it does not need other weather parameters as the case with the regression model. In general, (as shown in Fig. 9), both models can be used successfully for the estimation of $\overline{H}$ values throughout the year.

4. Conclusion

Results of this study show that the maximum values of monthly mean daily total and diffuse solar radiation in Mosul are 7.99 kWh/m^2 and 2.39 kWh/m^2 respectively, while the minimum values are 2.13 kWh/m^2 for total and 0.93 kWh/m^2 for diffuse solar radiation. Monthly mean hourly total and diffuse solar radiation are 958 Wh/m^2 and 269 Wh/m^2, respectively which reaches their maximum values between 13:00–14:00 h local time. The percentage distribution of clear days is quite high (57%), while for cloudy days is 11.5% and 31.1% for partly cloudy days in the year. It is deduced that hourly total solar radiation confirms Liu and Jordan's theoretical equation for estimating monthly mean hourly total solar radiation based on daily mean values. Several empirical equations are developed for estimation total and diffuse solar radiation. It is also shown that ARIMA(2,1,1) model can be used for prediction of daily clearness indices.

References

[1] Moon P. Proposed standard radiation curves for engineering use. J Franklin Inst 1940;230(5):583–617.
[2] Fritz S, Mcdonald TH. Average solar radiation in the United States. Heat Ventil 1949;46(7):61–4.
[3] Fritz S. Solar energy on clear and cloudy days. Sci Monthly 1957;84:55–65.
[4] Drummond AJ, Vowinckel E. The distribution of solar radiation throughout Southern Africa. J Meteorol 1957;14(4):343–53.
[5] Iqbal M. Correlation of average diffuse and beam radiation with hours of bright sunshine. Sol Energy 1979;23:169–73.
[6] Erbs D, Klein SA, Duffie JA. Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation. Sol Energy 1982;28(4):293–302.
[7] Abdalla YAG, Feregh GM. Contribution to the study of solar radiation in Abu Dhabi. Energy Convers Manag 1988;28(1):63–7.
[8] Islam MD, Alili AA, Kubo I, Ohadi M. Measurement of solar-energy (direct beam radiation) in Abu Dhabi, UAE. Renew Energy 2010;35(2):515–9.
[9] Islam MD, Kubo I, Ohadi M, Alili AA. Measurement of solar radiation in Abu Dhabi, UAE. Appl Energy 2009;86:511–5.
[10] Al-Riahi M, Al-Jumaily KJ, Ali HZ. Modeling clear-weather day solar-irradiance in Baghdad, Iraq. Energy Convers Manag 1998;39(12):1289–94.
[11] Boland J. Time series and statistical modeling of solar radiation. In: Badescu Viorel, editor. Recent advances in solar radiation modelling. Springer-Verlag; 2008. pp. 283–312.
[12] Huang J, Korolkiewicz M, Agrawal M, Boland J. Forecasting solar radiation on an hourly time scale using a Coupled Auto Regressive and Dynamical System (CARDS) model. Sol Energy 2013;87:136–49.
[13] Hassan A, Hejase N, Ali HA. Time-series regression model for prediction of mean daily global solar radiation in Al-Ain, UAE. ISRN Renew Energy; 2012. http://dx.doi.org/10.5402/2012/412471.
[14] Sulaiman MY, Hlaing oo WM, Wahab MA, Sulaiman ZA. Analysis of residuals in daily solar radiation time series. Renew Energy 1997;11(1):97–105.
[15] Zaharim A, Razali AM, Gim TP, Sopian K. Time series analysis of solar radiation data in the tropics. Eur J Sci Res 2009;25(4):672–8.
[16] Zeroual A, Ankrim M, Wilkinson A. Stochastic modeling of daily global solar radiation measured in Marrakesh, Morocco. Renew Energy 1995;6(7):787–93.
[17] Prescott J. Evaporation from a water surface in relation to solar radiation. Trans R Soc Sci Austr 1940;64:114–8.
[18] Angstrom A. Solar and terrestrial radiation. Report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation. Q J Roy Meteor Soc 1924;50(210):121–6.
[19] Iqbal M. An introduction to solar radiation. 1st ed. New York: Academic Press; 1983.
[20] Liu BYH, Jordan RC. The inter-relationship and characteristic distribution of direct, diffuse and total solar radiation. Sol Energy 1960;4(3):1–19.
[21] OriginLab, Northampton, MA (www.originlab.com).
[22] Nash J, Sutcliffe JV. River flow forecasting through conceptual models part I–a discussion of principles. J Hydrol 1970;10:282–90.
[23] Box G, Jenkins G, Reinsel G. Time series analysis: forecasting and control. 3rd ed. NJ, USA: Prentice Hall, Englewood Cliffs; 1994.
[24] Enders W. Applied econometric time series. 3rd ed. New York, NY, USA: John Wiley & Sons; 2010.
[25] Minitab 14 statistical software. State College, PA: Minitab, Inc; 2008.www.minitab.com.
[26] Lutkepohl H, Kratzig M. Applied time series econometrics. Cambridge, UK: Cambridge University Press; 2004 [JMulTi].
[27] Ljung G, Box G. On a measure of lack of fit in time series models. Biometrika 1978;65(2):297–303.
[28] Kang BO, Tam KS. A new characterization and classification method for daily sky conditions based on ground-based solar irradiance measurement data. Sol Energy 2013;94:102–18.
[29] Rassol RA. [M.Sc. thesis]. Baghdad-Iraq: Univ. of Technology; 1986.
[30] Aguiar RJ, Collares-Perrira M, Conde JP. Simple procedure for generating sequences of daily radiation values using library of Markov transition matrices. Sol Energy 1988;40(3):269–79.
[31] Mustacchi C, Cena V, Rocchi M. Stochastic simulation of hourly radiation sequences. Sol Energy 1979;23:47.
[32] Amato U, Andertta A, Bartolli B, Coluzzi B, Cuomo V, Fontana F, et al. Markov process and Fourier an analysis a tool to describe and simulate solar irradiance. Sol Energy 1986;37:179.