Modeling of Solar Radiation in the Time Series Domain (December 2006)

www.serd.ait.ac.th/reric

M.A. Alghoul*, M.Y. Sulaiman, B.Z. Azmi and M. Abd. Wahab

Abstract - *Time series analysis of relative global solar radiation from four stations in the Kingdom of Jordan, namely Amman, Aqaba, Dier Alla and Irbid have been undertaken. The deterministic component is removed by double differencing. The autocorrelation and partial autocorrelation functions of the residuals indicate autoregressive (AR) process. The coefficients of the autoregressive process and statistical indicators are estimated using the program PEST. The autoregressive representation of the residual series is justified by analyzing the autocorrelation and partial autocorrelation functions of the resulting white noise. To confirm that the residuals are observed values of independent and identically distributed random variables, Ljung-Box and McLeod-Li statistical test are applied to the residual series.*

Keywords - **Global solar radiation, Time series analysis, Kingdom of Jordan.**

1. INTRODUCTION

Time series analysis of solar radiation data is useful in predicting long-term average performance of solar energy system. Once a time series model is deduced for a set of regularly recorded observations, it can be used to generate future values as long as the statistical properties of the data remain the same. For solar radiation data this condition will always be satisfied.

Time series models find many applications in the field of engineering, science, sociology, economics, etc. It can be used to simplify description of data. Other applications of time series include, (i) separation (or filtering) of noise from signals, (ii) prediction of future values, (iii) testing hypothesis, (iv) prediction of time series from observation of another and (v) simulation studies.

In this paper, we apply the theory of time series to daily global solar radiation data from four stations in the Kingdom of Jordan. Prediction of future values is useful when a solar energy system is to be set up and its long-term performance needs to be evaluated. In the next section the formalism of time series is presented. The data is first removed of deterministic components using differencing technique. The residuals are subject to an algorithm to determine the order of the process. Both statistical and graphical analyses are applied to verify the dependent property of the series.

2. PREVIOUS WORK

Much work has been dedicated to the mathematical representation of the meteorological parameters as an important application in applied science. The data that were treated may consist of hourly or daily events.

M.A. Alghoul*, M.Y.Sulaiman, B.Z.Azmi and M. Abd. Wahab are attached to the Physics department, University Putra Malaysia, 43400 Serdang, Selangor, Malaysia.

*Corresponding author email:dr.alghoul@gmail.com

Time series of 20 years of daily solar radiation data from four Italian stations were analyzed on a statistical basic by Amato et. al (1986). It was shown that the radiation sequence was not stationary. The stochastic component followed a first order Markov model.

Boland (1995) devised a method to identify important cyclical components in solar radiation and ambient temperature data. After the contributions of the steady periodic part were removed the residuals of the time series were analyzed. It was shown that the daily solar radiation residuals were a first order autoregressive process while the daily average ambient temperature residuals were a third order autoregressive process.

Box-Jenkins approach was applied to daily solar radiation from four different locations in Malaysia by Sulaiman et al. (1997). The deterministic annual component was obtained using Fourier analysis. The stochastic component of the time series were fitted to three models i.e AR(1), AR(2), ARMA(1,1).

3. FORMALISM OF TIME SERIES PROCESS

The formalism of time series process can be found in a number of publications such as Box et al (1994), Fuller (1996), Kendal et al (1990), Pandit et. al (1983). In this work however we will follow the approach of Brockwell (1996).

A time series model for the observed data $\{x_t, t=0,\pm1,\ldots.\}$ is a specification of the joint distributions of a sequence of random variables $\{X_t, t=0,\pm1,\ldots.\}$ of which $\{x_t\}$ is postulated to be a realization.

For the time series $\{X_t\}$ the mean function is given as:

$$\mu_x(t) = E(X_t) = \sum_j^{\infty} X_{tj} P(X_{tj}) \tag{1}$$

where P is a probability distribution function of discrete random variables.

A covariance function of $\{X_t\}$ is given as:

$$\gamma_X(r,s) = Cov(X_r, X_s) = E[(X_r - \mu_x(r))(X_s - \mu_s(s))] \tag{2}$$

for integers r, s, and t.

It is assumed that the time series $\{X_t\}$ shall have finite second moment.

Loosely speaking, a time series $\{X_t\}$ is said to be stationary if it has statistical properties similar to those of the 'time-shifted' series $\{X_{t+h}\}$ for each integer h. It can be shown that (Brockwell, 1996) stationary time series obey the following conditions,

(i) $\mu_x(t)$ is independent of t

(ii) $\gamma_x(t+h,t)$ is independent of t for each h

If time series $\{X_t\}$ is stationary, an autocovariance function (ACVF) is defined as,

$$\gamma_X(t+h,t) = \mathrm{Cov}(X_{t+h}, X_t) \equiv \gamma_x(h) \tag{3}$$

where γ_x depends only on the index h known as the lag.

Also, the autocorrelation function (ACF) of a stationary time series $\{X_t\}$ is defined as,

$$\rho_x(h) = \gamma_x(h) / \gamma_x(0) = \mathrm{Cor}(X_{t+h}, X_t) \tag{4}$$

The ACVF and ACF provide a useful measure of the degree of dependences between the values of a time series at different times and for this reason play an important role in the prediction of future values using past and present values.

4. THE ARMA PROCESS

If a stationary time series $\{X_t\}$ satisfies the following equation,

$$X_t - \phi_1 X_{t-1} - .... - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + ... + Z_t + \theta_1 Z_{t-q} \tag{5}$$

where $\{Z_t\}$ is a sequence of uncorrelated random variables each with mean zero and finite variance, then it is an autoregressive moving average process of order p and q denoted by ARMA(p,q). The sequence of uncorrelated random variables is referred to as white noise denoted by WN(0,σ^2) having mean zero and variance σ^2. One important type of white noise is the independent and identically distributed random sequence (Brockwell, 1996). Independent and identically distributed noise plays an important role as a building block for more complicated time series models.

Equation (5) can be written in terms of the backward shift operator B defined by:

$$BZ_t = Z_{t-1} \tag{6}$$

In general,

$$B^m Z_t = Z_{t-m} \qquad m = 1,2.... \tag{7}$$

Therefore, equation (5) can be written as,

$$(1-\phi_1 B - \phi_2 B^2 - .... - \phi_p B^p) X_t = (1+\theta_1 B + \theta_2 B^2 + ... + \theta_q B^q) Z_t \tag{8}$$

or

$$\phi(B) X_t = \theta(B) Z_t \tag{9}$$

where,

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - .... - \phi_p B^p \tag{10}$$

and

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + ... + \theta_q B^q \tag{11}$$

The time series $\{X_t\}$ is said to be an autoregressive process of order p or AR(p) if θ (B) is equal to one and a moving average process of order q or MA(q) if ϕ (B) is equal to one.

The stationarity of the ARMA(p,q) process $\{X_t\}$ is ensured when

$$\phi(B) \neq 0 \ \textit{for all} \ |B| = 1 \tag{12}$$

In other words the zeros of the autoregressive polynomial must all be greater than one in absolute value.

5. THE SAMPLE ACF AND PACF

In practical problems we do not start with a model but with observed data $\{x_1, x_2, \ldots, x_n\}$. To assess the degree of dependence in the data and to select a model for the data that reflects this, the sample autocorrelation (sample ACF) and partial autocorrelation (sample PACF) functions are used. If the data are realized values of a stationary time series, then the sample ACF and PACF will provide with estimates of the ACF and PACF of the model time series $\{X_t\}$. These estimates may suggest which model is suitable for representing the dependence in the data.

For the sample time series $\{x_t$, t=1,2,….n$\}$

i) the sample mean is

$$\bar{x} = \frac{1}{n} \sum_{t=1}^{n} x_t \tag{13}$$

ii) the sample autocovariance function is

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}) \qquad -n < h < n \tag{14}$$

iii) the sample autocorrelation function ACF is

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} \qquad -n < h < n \tag{15}$$

iv) the sample partial autocorrelation function PACF is

$$\alpha(n) = \phi_{nn} = \frac{1}{\upsilon_{n-1}} \left[\hat{\gamma}(n) - \sum_{j=1}^{n-1} \phi_{n-1,j} \hat{\gamma}(n-j) \right] \tag{16}$$

with

$$\upsilon_n = \upsilon_{n-1} \left[1 - \phi_{nn}^2 \right] \tag{17}$$

and

$$\alpha(0) = 1 \tag{18}$$

$$\upsilon(0) = \bar{\gamma}(0) \tag{19}$$

$$\phi_{11} = \frac{\bar{\gamma}(1)}{\bar{\gamma}(0)} \tag{20}$$

The ACF and PACF are used to estimate the type and order of the process involved. As a rough guide, if the sample ACF falls between the plotted bounds $1.96/\sqrt{n}$ for lags $h > q$ then an MA(q) model is suggested, while if the sample PACF falls between the plotted bounds $1.96/\sqrt{n}$ for lags $h > p$, then an AR(p) model is suggested.

6. DETERMINISTIC COMPONENTS

The treatment thus far has assumed that the time series is stationary. In reality observed data are often made up of deterministic and random components. The deterministic observation may be the results of trend and periodic contributions. A plot of the sample data will reveal the deterministic components. For sample with substantial periodic component the autocorrelation function will also exhibit similar behavior with the same periodicity.

To subject the data to time series analysis, the deterministic component must be removed giving rise to the residuals. There are a number of methods of extracting the deterministic components. Sulaiman et al (1997) approximated the component with a Fourier series. In this work we use the method of Box and Jenkins (1994) where a differencing operator is applied to the original observed data until the differenced observations resemble a realization of some stationary time series

7. ELIMINATION BY DIFFERENCING

Let $\{y_t\}$ be the sample series containing the deterministic components. We define the lag-1 difference operator ∇ by:

$$\nabla y_t = y_t - y_{t-1} = (1 - B) y_t \tag{21}$$

where B is the backward shift operator defined previously. A lag-j difference can be written as:

$$\nabla_j y_t = y_t - y_{t-j} = (1 - B^j) y_t \tag{22}$$

Observed data that are effectively differenced produce residuals, thus,

$$x_t = y_t - y_{t-j} \tag{23}$$

8. GOODNESS OF FIT

If there is no dependence between the residuals then they can be regarded as observation of independent random variables and there is no further modeling to be done except to estimate their mean and variance. However, if there is significant dependence between the residuals, then a more complex stationary time series model must be found to account for the dependence. There exist simple tests to check the hypothesis that the residuals are observed values of independent and identically distributed random variables.

The sample autocorrelation function

About 95% of the sample autocorrelations of an independent and identically distributed random variables should fall between the bounds of $\pm 1.96/\sqrt{n}$ where n is the total number of data points.

Ljung and Box test (Ljung and Box, 1978)

In the Ljung and Box test use is made of a statistic defined as follows:

$$Q_{LB} = n(n + 2) \sum_{j=1}^{h} \hat{\rho}(j)/(n - j) \tag{24}$$

If $Q_{LB} > \chi^2_{1-\alpha}(h)$ where $\chi^2_{1-\alpha}(h)$ is the 1-α quantile of the chi-squared distribution with h degrees of freedom the independent and identically distributed random data hypothesis is rejected.

McLeod and Li test (McLeod and Li, 1983)

The McLeod and Li statistics is defined as follows:

$$\hat{Q} = n(n + 2) \sum_{k=1}^{h} \hat{\rho}^2_{ww}(k)/(n - k) \tag{25}$$

where $\hat{\rho}_{ww}(h)$ is the sample autocorrelations of the squared data. The hypothesis of the independent and identically distributed normal data is then rejected at level α if the observed value of $\hat{Q}$ is larger than the $(1-\alpha)$ quantile of the $\chi^2(h)$ distribution.

9. RESULTS AND DISCUSSION

The data is analyzed using a computer program PEST (Brockwell, 1996) using the formalism described earlier. The program can be used to plot, analyze and transform time series data. It can also be used to compute properties of time series models and fit models to data.

Table 1. Geographical information on the stations used

Stations	Latitude (N)		Longitude (E)		Elevation (m) (b.s.l)
	Deg	Min	Deg	Min	
Amman	31	59	35	59	772
Aqaba	29	33	35	00	51
Dier Alla	31	13	35	37	-224
Irbid	32	33	35	51	616

Table 2. Number of data points and differencing lags used

Stations	Total number of data points	Lag		Number of data points after differencing
		First differencing	Second differencing	
Amman	1826	366	192	1268
Aqaba	730	365	179	186
Dier Alla	366	180		186
Irbid	730	365	179	186

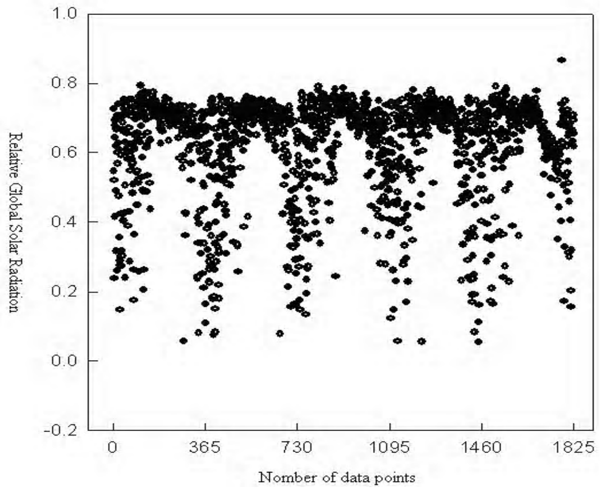

Fig. 1. Plot of relative global solar radiation for Amman

The solar radiation time series $\{y_t\}$ are derived from daily relative solar radiation data of four stations in the Kingdom of Jordan. Information on these stations are given Table 1. The relative data are obtained by taking the ratio of the mean daily solar radiation data and the daily extraterrestrial radiation given by Duffie and Beckman (1991). The total number of data points used in the analysis is given in Table 2. In Fig. 1, the plot of the relative global solar radiation data against the number of data points for Amman is shown. The plot clearly reveals the deterministic components in a form of periodic variations. A differencing analysis is then applied to the data with lag equals to the approximate period of the seasonal components. The lags used in the differencing analysis are given in Table 2. The series of the differenced data of all the stations are again plotted to determine the outcome. It is found that, except for Dier Alla not all of the deterministic components are satisfactorily removed. A second differencing is carried out with lags given in Table 2.

The ACFs and PACFs of the residuals are then determined using the PEST computer program. The ACF for Amman is plotted in Fig. 2 and the PACF in Fig. 3. Generally, there exists significant number of points that fall outside the bounds of $1.96/\sqrt{n}$. Thus, the residuals are not generated by a sequence of white noise.

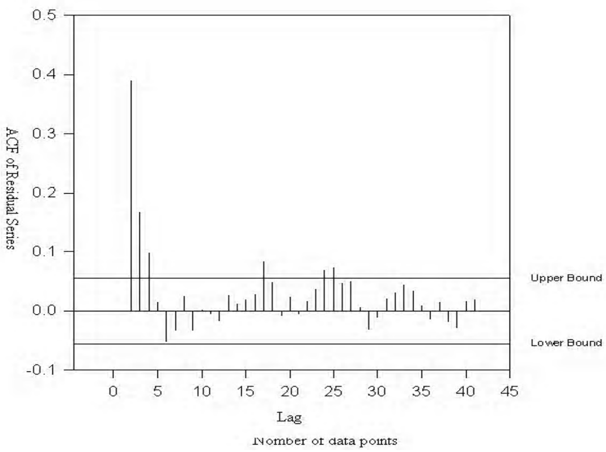

Fig. 2. Plot of the autocorrelation function of the residual for Amman

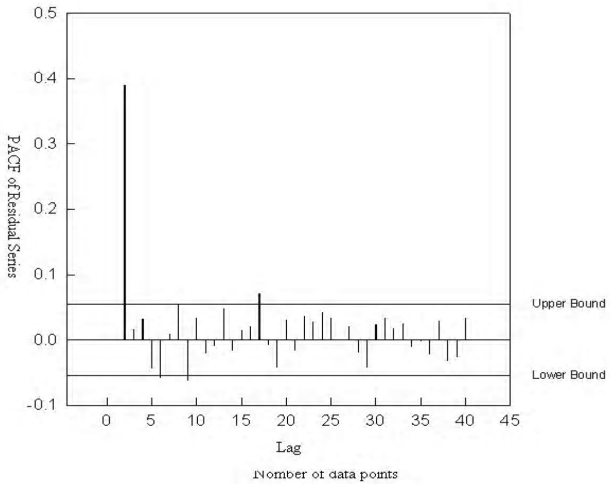

Fig. 3. Plot of the partial autocorrelation function of the residual for Amman

Table 3(a). Autoregressive coefficients for Amman

Order of coefficients, p	AR coefficients, θ_p	Ratio of oefficients to 1.96*standard error
1	0.395	7.01917
2	0.010	0.01642
3	0.046	0.77740
4	-0.026	-0.43500
5	-0.062	-0.98127
6	-0.103	-0.17530
7	0.073	1.30960
8	-0.065	-1.12690

Table 3(b). Autoregressive coefficients for Aqaba

Order of coefficients, p	AR coefficients, θ_p	Ratio of coefficients to 1.96*standard error
1	0.422	3.00990
2	-0.015	-0.09694
3	-0.020	-0.13204
4	0.022	0.140030
5	0.088	0.57270
6	-0.057	-0.37379
7	0.038	0.24784
8	0.018	0.12018
9	-0.066	-0.43500
10	-0.095	-0.62380
11	0.067	0.44005
12	-0.057	-0.37374
13	-0.036	-0.23158
14	0.031	0.20385
15	-0.019	-0.12300
16	-0.074	-0.48349
17	0.120	0.78456
18	-0.058	-0.37914
19	0.074	0.47928
20	0.049	0.32207
21	-0.126	-0.82136
22	-0.019	-0.12281
23	-0.003	-0.02259
24	0.004	0.02741
25	-0.159	-1.22300

Table 3(c). Autoregressive coefficients for Dier Alla

Order of coefficients, p	AR coefficients, θ_p	Ratio of coefficients to 1.96*standard deviation
1	0.433	3.0969
2	0.214	1.5353

Table 3(d). Autoregressive coefficients for Irbid

Order of coefficients, p	AR coefficients, θ_p	Ratio of coefficients to 1.96*standard error
1	0.392	2.6880
2	0.067	0.0449
3	0.029	0.1957
4	-0.038	-0.2539
5	-0.047	-0.3137
6	-0.021	-0.1406
7	0.024	0.1636
8	0.109	0.7290
9	-0.046	-0.3086
10	0.012	0.0840
11	-0.071	-0.4741
12	0.079	0.5284
13	-0.018	-0.1192
14	0.077	0.5112
15	-0.060	-0.3967
16	0.062	0.4161
17	0.0627	0.4172
18	-0.081	-0.5369
19	0.064	0.4278
20	-0.050	-0.3325
21	0.017	0.1152
22	0.019	0.1280
23	0.003	0.0217
24	0.105	0.7002
25	-0.096	-0.6360
26	0.202	1.3498

A model of the residual is then estimated using PEST program. The coefficients of the AR(p) model and the variance of the white noise are determined. The ratios between the values of the parameters and 1.96 times the standard error are also calculated. If these ratios are greater than one in absolute value than it can be concluded that at level of 0.05 the corresponding coefficients are different from zero. The output of the PEST estimation are given in Tables 3(a), 3(b), 3(c), 3(d). Based on these estimations, the relative global solar radiation data can be described by the following processes:

$X_t=Z_t+0.395X_{t-1}-0.062X_{t-5}+0.073X_{t-7}-0.065X_{t-8}$ for Amman

$X_t=Z_t+0.422X_{t-1}-0.159X_{t-25}$ for Aqaba

$X_t=Z_t+0.433X_{t-1}+0.214X_{t-2}$ for Dier Alla

$X_t=Z_t+0.392X_{t-1}+0.202X_{t-26}$ for Irbid

The ACFs and PACFs of the model are then calculated together with the random test statistics. The ACF for Amman is shown in Fig. 4 and the PACF in Fig. 5. In Table 4, we list the test statistics for all the stations used in the analysis. The ACFs and PACFs indicate white noise property with almost all points falling within the bounds. Also, all the test statistics are found to be less than the values of the chi-squared for 40 degrees of freedom. The values of the white noise variance are given in table 5 for all the stations. Thus, the AR(p) models describe well the relative global solar radiation data for all the four stations in the Kingdom of Jordan.

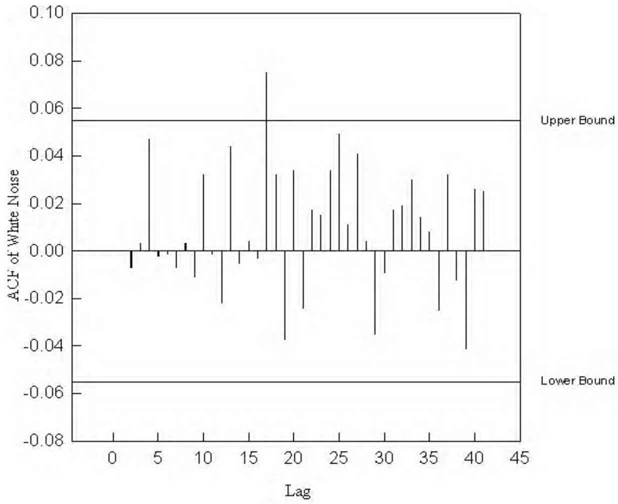

Fig. 4. Plot of the autocorrelation function of the white noise for Amman

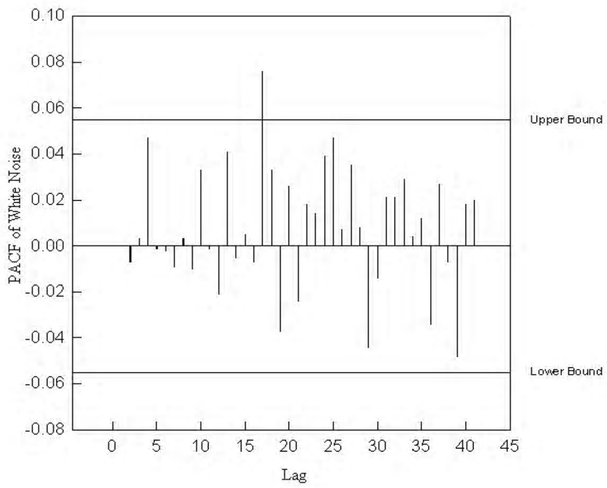

Fig. 5. Plot of the partial autocorrelation function of the white noise for Amman

Table 4. Random test statistics

Stations	Q_{LB}(h=40)	$\hat{Q}$ (h=40)	χ^2(h=40)
Amman	40.1	53.4	55.8
Aqaba	23.2	49.6	55.8
Dier Alla	47.8	25.8	55.8
Irbid	24.2	48.1	55.8

Table 5. White noise variances

Stations	White noise variance, σ^2
Amman	0.0537
Aqaba	0.0228
Dier Alla	0.0323
Irbid	0.0446

10. CONCLUSIONS

Relative global solar radiations of four stations in the Kingdom of Jordan are analyzed using time series technique. The deterministic components are removed by double differencing. The ACFs and PAC's of the sample indicate data dependency. The sample is fitted to AR(p) models and the ACFs, PACFs and random test statistics of the models are evaluated. These parameters support AR(p) representation of the data.

ACKNOWLEDGEMENTS

The work is financially supported by the Ministry of Science, Technology and Environment, Malaysia under the IRPA (Intensification of Research in Priority Areas). We hereby wish to acknowledge the financial assistance of the government of Malaysia.

NOMENCLATURE

B	backward shift operator
h	time series lag
p	order of autoregressive process
q	order of moving average process
Q_{LB}	Ljung-Box statistic
$\widehat{Q}$	McLeod-Li statistic
t	time series index
$\overline{x}$	sample mean function
x_t	time series of observed data
X_t	model time series
Z_t	white noise
y_t	sample time series containing deterministic components

Greek letters

μ	mean function of time series
Γ	covariance function of a time series
$\hat{\gamma}$	sample autocovariance function
ρ	autocorrelation function
$\hat{\rho}$	sample autocorrelation function
α	sample partial autocorrelation function
ϕ	autoregressive coefficients
θ	moving average coefficients

REFERENCES

[1] Amato, U., Andretta, A., Bartoli, B., Colussi, B., Cuomo, V., Fontana, F., and Serio, C. 1986. Markov process and Fourier Analysis as a Tool to Describe and Simulate Daily Solar Radiation, Solar Energy, 37, 179-194.

[2] Boland, John. 1995. Time Series Analysis of Climatic Variables, Solar Energy, 55, 377-388.

[3] Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. 1994. Time Series Analysis: Forecasting and Control, 3rd edn, Prentice-Hall, N. J.

[4] Brockwell, P. J. and Davis, r. A. 1996. Introduction to Time Series and Forecasting, Springer-Verlag, New York.

[5] Duffie, J, and Beckman, W. 1991. Solar Engineering of Thermal Processes, 2nd edn, John Wiley, New York.

[6] Fuller, W. A. 1996 Introduction to Statistical Time Series, 2nd edn, John Wiley & Sons, Inc. New York.

[7] Kendall, S. M. and Ord. J. K. 1990. Time Series, 3rd edn, John Wiley & Sons Inc. New York.

[8] Ljung, G. M. and Box, G. E. P. 1978. On Measure Lack of Fit in Time Series Model, Biometrika, 65, 297-304.

[9] McLeod, A. I. and Li, W. K. 1983. Diagnostic Checking ARMA Time Series Models Using Squared Residual Autocorrelations, J. Time Series Anal., 4, 269-273.

[10] Pandit, S. M. and Wu, S. M. 1983. Time Series and System Analysis with Applications, John Wiley & Sons Inc. New York.

[11] Sulaiman Yusof M., Hlaing Oo W. M., Mahdi Abd. Wahab and Sulaiman, Z. A. 1997. Analysis of Residuals in Solar Radiation Time Series, Renewable Energy, 11, 97-105.