We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



8

MATLAB-Assisted Regression Modeling of Mean Daily Global Solar Radiation in Al-Ain, UAE

Hassan A. N. Hejase and Ali H. Assi
United Arab Emirates University
United Arab Emirates

1. Introduction

Many researchers have modeled weather data using classical regression, time-series regression and Artificial Neural Networks (ANN) techniques. MATLAB is used in handling data and writing task specific codes for models as well as in performing statistical analysis and curve fitting works. This is due to the dynamic nature of MATLAB and its rich toolboxes that cover almost every aspect of mathematical and statistical engineering applications.
Numerous authors (Abdalla & Feregh, 1988; Assi & Jama, 2010; Akinoglu & Ecevit, 1990; Al Mahdi et al., 1992; Ampratwum & Drovlo, 1999; Elagib & Mansell, 2000; Khalil & Alnajjar, 1995; Menges et al., 2006; Newland, 1988; Podesta' et al., 2004; Sahin, 2007; Samuel, 1991; Ulgen & Hepbasli, 2002) to count few developed empirical regression models to predict the monthly average daily global solar radiation (GSR) in their region using various parameters. The mean daily sunshine duration was the most commonly used and available parameter. The most popular model was the linear model by Angström-Prescott (Podesta' et al., 2004; Assi & Jama, 2010) which establishes a linear relationship between GSR and sunshine duration with knowledge of extra-terrestrial solar radiation and the theoretical maximum daily solar hours. Many studies with empirical regression models were done for diverse regions around the world. (Menges et al. 2006) reviewed 50 GSR empirical models available in literature for computing the monthly average daily GSR on a horizontal surface. They tested the models on data recorded in Konya, Turkey for comparison of model accuracy. The number of weather parameters varied between models. The diverse regression models used include linear, logarithmic, quadratic, third order polynomial, logarithmic-linear and exponential and power models relating the normalized GSR to normalized sunshine hours. Other models included in Menges work used direct regression models involving various weather parameters such as precipitation, cloud cover, etc., in addition to geographical data (altitude, latitude). (Şahin, 2007) presented a novel method for estimating the solar irradiation and sunshine duration by incorporating the atmospheric effects due to extra-terrestrial solar irradiation and length of day. The author compares his model with Angström's equation with favourable advantages as his method does not use Least Square Method in addition to having no procedural restrictions or assumptions. (Ulgen & Hepbasli, 2002) developed two empirical correlations to estimate the monthly average daily GSR on a horizontal surface for Izmir, Turkey. Their models resemble Angström type equations. They

compared their models with 25 models previously reported in literature on the basis of statistical error tests (MBE (mean bias error), RMSE (root mean square error), MPE (mean percentage error), and R^2 (coefficient of determination)) with favourable results.

Other authors worked on prediction models based on artificial neural network techniques and most specifically Multi-Layer Perceptron (MLP) and Radial-Basis Function (RBF) methods (Al-Alawi & Al-Hinai, 1998; Assi et al., 2010; Behrang et al., 2010; Benghanem et al., 2009; Boccol et al., 2010; Elminir et al., 2005; Krishnaiah et al., 2007; Mohandes et al., 1998; Mohandes et al., 2000; Rehman & Mohandes, 2008). The advantage of the ANN models is their ability to handle large amounts of data as well as the ability to handle random data without worry of incomplete, inaccurate or noise-contaminated data.

Another approach followed by many researchers is using time-series modeling techniques which employ regression for the deterministic component and Box-Jenkins Auto-Regressive Integrated Moving Average (ARIMA) modeling for the stochastic residual component (Box & Jenkins, 1994; Enders, 2010). (Sulaiman et al., 1997) and (Zaharim et al., 2009) use the ARMA Box-Jenkins method to model GSR data in Malaysia. They model the data using non-seasonal autoregressive models where the model adequacy is checked using the Ljung-Box statistic for diagnostic data. However, only short-term data was used for testing their model. (Reikard, 2009) employs a combination of logarithmic regression and ARIMA modeling to predict solar radiation at high resolution and compares his models with other forecast methods such as ANN and considered the 24-hour daily seasonality not taken into account in most modeling approaches. (Boland, 1995) and (Zeroual et al., 1995) use a combination of regression model with Fourier series for the deterministic part and ARMA modeling for the residual stochastic component.

The potential of solar energy harvesting in the UAE is significant, with an average annual sunshine hours of 3568 h (i.e. 9.7 h/day), which corresponds to an average annual solar radiation of approximately 2285 kWh/m^2, i.e. 6.3 kWh/m^2 per day (Assi & Abdi, 2010).The main limitation of using many weather parameters is the difficulty of obtaining this data due to the high expenses and availability of recording equipment. Some of the regression models are more accurate for monthly data than daily data and most published works only show monthly data comparison results as the daily mean data are less accurate.

The authors are doing extensive work on the analysis of weather data in three UAE cities, namely, Al-Ain, Abu Dhabi and Sharjah. The prediction techniques employed range from classical one-parameter based regression techniques (Sunshine data), Artificial Neural Networks (ANN) Multi-Layer Perceptron (MLP) and Radial-Basis Function (RBF) techniques, and lately using time-series regression techniques with Auto-Regressive Integrated Moving-Average (ARIMA) modeling. Most of the work done relies on the use of MATLAB as a common platform for preparing and processing data as well as for testing the suggested prediction models. MATLAB was also used in curve fitting the output data and finding the statistical error parameters that judge the accuracy and dependability of our prediction models. The used correlations included the linear Angstrom-Prescott model and its derivations, namely, the second and third order correlations, in addition to the single term exponential model, logarithmic model, linear logarithmic model and power model.

This chapter uses examples with classical empirical regression techniques and time-series techniques to predict the monthly average daily GSR data in Al-Ain, UAE. In time-series regression technique, the deterministic component was modeled by decomposing it into a multiple linear regression component as a function of the available weather variables plus a

seasonal component accounting for the annual periodicity and a linear trend. The residual error is studied using Box-Jenkins ARIMA modeling techniques for the sake of further enhancing the predicted solar radiation data. The resultant noise residual error renders Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) plots within the 95 % confidence interval bounds and a quasi-normal noise error with zero mean and constant variance. The stationary form of the resulting time-series and the white-noise type residual error provide extra confidence of the long-term prediction accuracy of the estimated model. The MATLAB codes written by authors and their group involve employing various data modeling techniques such as linear regression, curve fitting, de-trending, FFT algorithm, statistical data analysis as well as Box-Jenkins method to predict a time-series model for the residual error.
The current work on solar radiation data in Al-Ain City, UAE will be correlated with other ongoing studies by authors to come up with solar radiation prediction models for the UAE cities of Abu Dhabi and Sharjah. The final objective is to come up with a good national weather model capable of predicting the mean monthly GSR for the whole UAE within an acceptable prediction error.

2. Methodology

The weather database meteorological data provided by the National Center of Meteorology & Seismology (NCMS) in Abu Dhabi) for the periods between 1995 and 2007 was divided into two sets: A model data set with daily record of the variables: air temperature, wind speed, sunshine hours and relative humidity for the years 1995- 2004 (10 years), and a test data set for the years of 2005-2007. All regression modeling and simulation work is done using MATLAB tools and with the help of the statistical software packages Minitab (Minitab, 2010) and SPSS (IBM SPSS Statistics, 2010). The next section will explain the modeling procedure for both the classical empirical regression and time-series regression techniques. Validation of model accuracy is presented along with corresponding error statistics. Then, a comparison between both empirical approaches is made.

2.1 Procedure

In this section we discuss two methods used to generate the weather data models for Al-Ain, UAE for years 1995-2004. Selected models are validated with data from years 2005-2007. The two models discussed in this chapter are:

1. Classical regression methods (Empirical models)
2. Time-series regression model (Regression with Box-Jenkins ARIMA method)

The use of MATLAB in each approach will be fully discussed along with the details of commands used and computation results for the models under investigation.

2.1.1 Classical regression modelling approach (empirical models)

The empirical regression models are generated with help of MATLAB and verified using SPSS. The appropriate MATLAB commands used are addressed with each procedural step.

2.1.1.1 Data preparation for nonlinear regression models

The 10-year mean daily GSR data in Al-Ain, UAE is computed for both sunshine hours and mean daily GSR yielding a data set of size N=365 for each variable. In MATLAB, this is done as follows:

```
N= 365;   NY= 10;
% N= 365 days of the year and NY= # years of model data used (1995-2004)
for k=1:N,      % Day index of year
   GSR(k)= mean( GSR_10yr (k:N:(NY-1)*N+k));
end
```

where GSR_10yr is the data sheet containing the daily mean GSR data array of size 3650 days (1995-2004). Similarly, we generate the 10-year average daily SSH from the sunshine hours data.

2.1.1.2 Extra-terrestrial radiation parameters

Mean daily values of GSR data are calculated from the knowledge of the latitude and longitude in the city of Al-Ain (Latitude = $24^0 16'$ and Longitude = $55^0 36'$). The extraterrestrial solar radiation on horizontal surface in kWh/m^2 (G_0) and theoretical maximum daily sun hours (S_0) are calculated from the equations (Assi et al., 2010):

$$G_0 = \frac{24 G_{SC}}{\pi}\left[1 + 0.033\cos\left(\frac{360n}{365}\right)\right]\left[\cos(\phi)\cos(\delta)\sin(\omega_s) + \omega_s \sin(\phi)\sin(\delta)\right] \quad (1)$$

$$S_0 = \frac{2}{15}\omega_s \quad (2)$$

where n is the day index, ω_s the mean sunrise hour angle for the month, ϕ the latitude, and δ the declination angle. G_{SC} is a constant representing the daily extraterrestrial solar radiation on horizontal and is given by 1.367 kWh/m^2. The declination angle (δ) is defined by the equation:

$$\delta = \sin^{-1}\left[\sin\left(\frac{23.45\pi}{180}\right)\sin\left(\frac{2\pi[n+284]}{365}\right)\right] \quad (3)$$

The GSR and SSH data are next normalized to the extraterrestrial values G_0 and S_0 described in eq. (1)-(2), and the resulting normalized data arrays denoted by clearness index (RSSH=GSR/G_0) and Sunshine duration ratio (RSSH = SSH/S_0) are stored in excel file solardata.xlsx in the form:

```
>> gsrdata=[N, S0, G0, RSSH, RGSR];
>> s=xlswrite('solardata.xlsx',gsrdata,'Sheet1', 'A1:E365');
```

The RGSR-RSSH data is then fitted to different nonlinear regression models as per Table 1.

Model reference	Type	Equation
(Podesta, 2004)	Linear	$y= b_1+b_2* x$
(Akinoglu & Ecevit, 1990)	Quadratic	$y= b_1+b_2* x+b_3* x^2$
(Samuel, 1991)	Cubic	$y= b_1+b_2* x+b_3* x^2+b_4* x^3$
(Ampratwum & Dorvlo, 1999)	Logarithmic	$y= b_1+ b_2 *\log(x)$
(Newland, 1988)	Log-Linear	$y= b_1+ b_2* x+ b_3*\log(x)$
(Elagib & Monsell, 2000)	Exponential	$y= b_1* \exp(b_2*x)$

Table 1. Nonlinear empirical regression models used for Al-Ain, UAE weather data

2.1.1.3 MATLAB application in nonlinear regression

The nonlinear regression in MATLAB is performed using two tools:

1. Interactive nonlinear regression toolbox (nlintool)
2. Nonlinear mixed-effects estimation (nlmefit)

The procedure followed in each approach is discussed in the next sections including sample output results from the weather GSR modeling. Alternative approaches include specifically written MATLAB m-files or the use of commercial statistical software packages such as SPSS, Minitab or SAS. All our MATLAB results agree well with results obtained using SPSS and Minitab.

2.1.1.3.1 Use of the interactive nonlinear regression toolbox (nlintool)

The MATLAB procedure to use "nlintool" is explained through the application of the log-linear regression model, described in Table 1, to model the relation between the clearness index (RGSR) and sunshine duration ratio (RSSH) through the following steps:

1. Given RSSH and RGSR data (N points each). Read the input data to MATLAB workspace from an Excel data sheet "GSRdata.xlsx" as follows:

```
>> inputdata=xlsread ('GSRdata.xlsx', 'sheet1','A1:B365')
>> RSSH=inputdata(:,1);
>> RGSR=inputdata(:,2);
```

   The input data array contains two columns of length N=365, namely, RSSH (independent variable or predictor) and RGSR (dependent variable).
2. Next, generate an m-file script called "loglinear.m" containing the regression function as follows:

```
function yhat = loglinear(beta, x)
      b1 = beta(1);
      b2 = beta(2);
      b3 = beta(3);
yhat = b1 + b2*x + b3*log(x);
```

3. In command window type:

```
>> beta0=[ 0 0 0]; % define initial guess of constants beta(k), k=1,2,3
>> nlintool(RSSH, RGSR,@loglinear,beta0)  % Invoke the interactive
   nonlinear regression toolbox
```

The output is the log-linear regression curve with error bounds as shown in Fig. 1.

4. Data that can be downloaded to MATLAB workspace include the regression coefficients (beta) and root-mean square error (rmse). This is done as follows:

```
>> rmse   % root mean-square error
rmse =
  0.0262
>> beta   % regression model coefficients
beta=
 -0.1778  6.1287E-05   0.2251
```

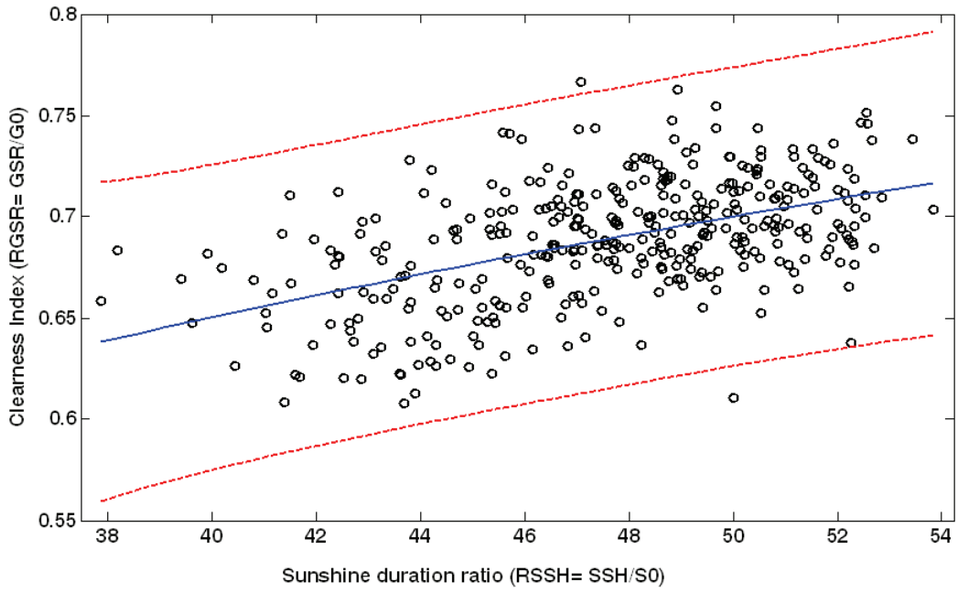


Fig. 1. Nonlinear regression curve with a log-linear function between Clearness index (RGSR) and sunshine duration ration (RSSH)

The command "whos" provides information on available workspace variables. Similarly, other nonlinear regression curve fits could be obtained by generating the corresponding m-file for each regression function and specifying the corresponding initial guess values for the unknown parameters.

Table 2 shows the coefficients obtained for each regression model shown in Table 1. The statistical error parameters resulting from each regression model can be computed using SPSS, Minitab or using MATLAB from the expressions displayed in section 2.3.

Type	Equation	b1	b2	b3	b4
Linear	y= b1+b2* x	0.4619	0.0047		
Quadratic	y= b1+b2* x+b3* x2	0.2980	0.0118	-7.400E-05	
Cubic	y= b1+b2* x+b3* x2+b4* x3	8.4297	-0.5183	0.0114	-8.2410E-05
Logarithmic	y= b1+b2*log(x)	-0.1697	0.2223		
Log-Linear	y= b1+ b2* x+ b3*log(x)	-0.1778	-6.133E-05	0.2252	
Exponential	y= b1* exp(b2*x)	0.4948	0.0069		

Table 2. Estimated coefficients for empirical regression models of weather data for years 1995-2004 in Al-Ain, UAE. Variables are y = RGSR = GSR/G0 and x = RSSH = SSH/S0

2.1.1.3.2 Use of the nonlinear mixed-effects estimation (nlmefit)

The nonlinear regression in MATLAB can also be performed using the nonlinear mixed-effects estimation (nlmefit). This tool fits the model by maximizing an approximation to the marginal likelihood with random effects integrated out, assuming that [See MATLAB help] random effects are multivariate normally distributed and independent between groups, and that observation errors are independent, identically normally distributed, and independent of the random effects. As an example, the format for the log-linear regression equation is as follows:

```
>> model = @(phi,t)(phi(:,1)+ phi(:,2).*t + phi(:,3).*log(t));
>> phi0 =[0 0 0];
>> group=[1:365];
>> [beta,psi,stats] = nlmefit(RSSH(:),RGSR(:),group,[],model,phi0)
```

The arguments of the MATLAB function 'nlmefit' used for our regression are:
RSSH: is an n-by-1 array of n observations on 1 predictor.
RGSR: is an n-by-1 vector of responses.
Group: is a grouping variable indicating m groups in the observations. Here, we enter size of rows of predictors, i.e. group=[1:365].
Model: is a function handle that accepts predictor values and model parameters and returns fitted values.
phi0: contains the initial values of the regression for equation parameters.
The program outputs are:
beta= contains the estimated regression model parameters
psi= an r-by-r estimated covariance matrix for the random effects. By default, r is equal to the number of model parameters p.
stats= returns a structure with the following parameters:

- logl — The maximized log-likelihood for the fitted model
- mse — The estimated error variance for the fitted model
- aic — The Akaike information criterion for the fitted model
- bic — The Bayesian information criterion for the fitted model
- sebeta — The standard errors for beta
- dfe — The error degrees of freedom for the model

The following is the MATLAB output to the log-linear regression model:

```
beta =
  -0.1778
  -6.1336E-05
   0.2252
psi =
  1.0e-003 *
   0.6821         0         0
        0    0.0000         0
        0         0    0.0000
stats =
logl: 8.1257e+002
mse: 3.7845e-028
rmse:1.9454e-014
aic: -1.6111e+003
bic: -1.5838e+003
sebeta: [1.3834 0.0104 0.4869]
dfe: 358
```

The regression model parameters obtained using the 'nlmefit' function match with the values computed using the "nlintool" function and listed in Table 2. The best regression

model should yield the lowest value of AIC, BIC and RMSE in addition to the autocorrelation test requirements discussed later in the chapter. Fig.2 shows the ten-year mean daily GSR computed from the log-linear regression model with estimated coefficients listed in Table 2. Note the very good agreement between regression model and measured data. The error statistics yield a deterministic coefficient of R^2= 97.74 % in addition to low statistics RMSE = 0.2104, MBE = -0.0755, MABE = 0.1872, and MAPE= 3.11 %. A comparison of the remaining regression yields excellent agreement with measured data with R^2 values exceeding 96 %.

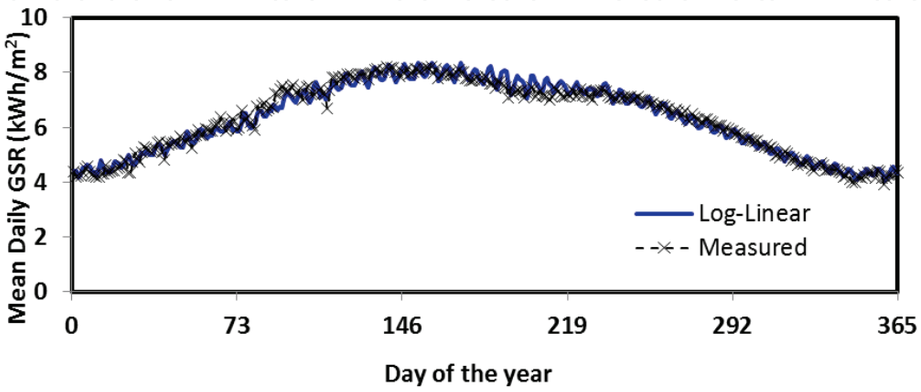


Fig. 2. Ten-year mean daily GSR data comparison between the empirical log-linear regression model and measured data in Al-Ain, UAE for years 1995-2004

2.1.1.4 Validation of empirical regression models

The six selected empirical regression models are validated by computing the predicted GSR data from these models using the test data set of years 2005-2007. The empirical models compare very well with measured data for the test data period as depicted in Fig. 3. All models yield deterministic coefficients values R^2 better than 98 %. The models also yield low values for RMSE, MBE, MABE and MAPE indicating their adequacy as weather prediction models for Al-Ain, UAE. The 3rd-order polynomial regression model (Cubic) outperforms the other five empirical models with lowest error statistics and highest deterministic coefficient.

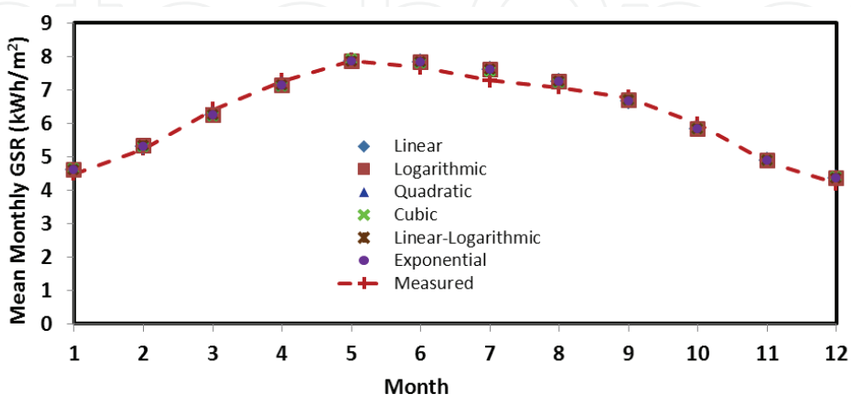


Fig. 3. Comparison of monthly mean regression model GSR with measured data for test period of 2005-2007

2.1.2 Time-series regression modelling with ARIMA approach

The time-series regression modelling approach makes use of the four weather parameters measured daily over a period of 10 years, i.e. 1995-2004 in Al-Ain, UAE. These parameters are:

1. Mean daily temperature (^{0}C) abbreviated as T
2. Mean daily wind speed (knots) abbreviated as W
3. Daily sunshine hours abbreviated as SSH
4. Mean daily relative humidity (%) abbreviated as RH
5. Mean daily global solar radiation (kWh/m2) abbreviated as GSR

The modeling procedure steps are as follows:

Step 1. Explore the form and characteristics of the dependent variable (GSR) and the predictors (T, W, SSH and RH) by looking at their time-series plots with the help of MATLAB. These plots will help in identifying the behavior of each variable and to check the trend and seasonality of GSR data.

Step 2. Compute the descriptive statistics of the four independent variables (T, W, SSH, and RH) and the dependent variable (mean daily GSR). The correlation between these variables is shown in Table 3. It shows that wind speed has the least effect on GSR performance. From the time-series plots of GSR, T and SSH data we can observe a cyclical behavior. Wind and humidity data on the other hand are more random. The best model is obtained when all four independent variables are considered in the GSR regression model.

	T	**W**	**SSH**	**RH**	**GSR**
T	1.0000	0.1907	0.4555	-0.7419	0.7958
W	0.1907	1.0000	-0.1099	-0.2170	0.1425
SSH	0.4555	-0.1099	1.0000	-0.4566	0.6972
RH	-0.7419	-0.2170	-0.4566	1.0000	-0.6761
GSR-kWh/m^2	0.7958	0.1425	0.6972	-0.6761	1.0000

Table 3. Correlation between the measured weather data parameters for the city of Al-Ain, UAE for years 1995-2004

The Pearson correlation values between the response (GSR) and the four predictor variables (T, W, SSH, RH) are found in MATLAB using the command:

```
>> corr(T, W, SSH, RH, GSR)
```

Table 3 shows that the temperature and sunshine hours have dominant effect on the GSR parameter followed closely by relative humidity and with less influence by wind speed.

The computation of the descriptive statistics can be done directly in SPSS or Minitab. In MATLAB we can either write a script m-file to compute all needed statistical parameters or we can use the already available functions in MATLAB and other functions that can be run under the **Statistics toolbox**. One MATLAB command that generates some statistical parameters is:

```
>> [xds,yds] = datastats (xdata, ydata)
```

which returns statistics for the column vectors xdata and ydata to the structures xdata and ydata, respectively. xdata and ydata must be of the same size. The returned statistics include: sample size, maximum value, minimum value, mean, median, range, and standard deviation. MATLAB also has standalone functions for statistical values such as: max, min, length (sample size), mean, median, mode, std (standard deviation), and var (variance). The coefficient of variation is found from the ratio of std(x)/mean(x). One can also use hist(x) to get a histogram of the sample data. Other statistical measures need to be programmed or can be determined using the MATLAB Statistics toolbox. Table 4 lists the commands that can be employed to obtain the central and dispersion measures for the data sets under study.

Central Measures	
Command	**Description**
geomean	Geometric mean
harmean	Harmonic Mean absolute deviation
mean	Arithmetic mean
median	50 % Percentile
trimmean	Trimmed mean
mode	Most frequent value

Dispersion Measures	
Command	**Description**
iqr	Interquartile range
mad	Mean absolute deviation
moment	Central moment of all orders
range	Range
std	Standard deviation
var	Variance

Table 4. MATLAB commands for computing central and dispersion measures of a data set

The following commands can be used under the Statistics toolbox:

```
>> locate = [geomean(x) harmmean(x) mean(x) median(x), trimmean(x,25)]
>> stats = [iqr(x) mad(x) range(x) std(x)]
```

On the other hand, measures of shape are found using quantiles ($0 < p < 1$) or percentiles ($0< p <100$). The percentiles for a data sequence xdata are found in the Statistics toolbox using the command:

```
>> y = prctile(xdata,p);   % p= percentile needed
>> y = quantile(xdata,p);   % p= quantileneeded
```

The shape of a data distribution is also measured by the Statistics Toolbox functions skewness, kurtosis, and, more generally, moment.
Table 5 shows the descriptive statistics obtained for the mean daily global solar radiation data for Al-Ain, UAE for period 1994-2005.

Step 3. Fig. 4 shows the time-series plot of the mean daily GSR for years 1995-2004 with leap days excluded. The plot shows a clear periodicity of one year (365 days).

Step 4. The partial least square regression technique is used to model the relation between GSR data for ten years (1995-2004) and the four aforementioned dependent weather variables. The model equation obtained using MATLAB or using either software packages SPSS or Minitab is of the form:

$$GSR_{regression}(t) = a_1 + a_2 * T(t) + a_3 * W(t) + a_4 * SSH(t) + a_5 * RH(t) \quad (4)$$

where the regression coefficients are given in Table 6.

Statistical Parameters	T	W	SSH	RH	GSR
Mean (μ)	36.501	7.5305	9.9472	44.0450	6.3332
StdDev (σ)	7.7250	1.9896	1.5849	16.8160	1.3962
SE Mean($\sigma/\sqrt{N}$)	0.1280	0.0329	0.2620	0.2780	0.0231
Variance (σ^2)	59.6690	3.9585	2.5121	282.7660	1.9494
CoefVar (σ/μ) in %	21.1600	26.4200	15.9300	38.1800	22.0500
Min	16.2000	0.2084	0.8000	7.3750	1.1350
Median	37.6000	7.2506	10.2000	43.4170	6.5540
Max	49.1000	22.7936	12.8000	97.2300	8.8210
Skewness (σ_k)	-0.3000	1.6000	-1.5200	0.2000	-0.3600
Kurtosis (ku)	-1.2100	5.7300	3.7700	-0.7000	-0.9000

Table 5. Descriptive data statistics for the measured weather data parameters for the city of Al-Ain (Years 1995-2004) where N= 3650 data samples

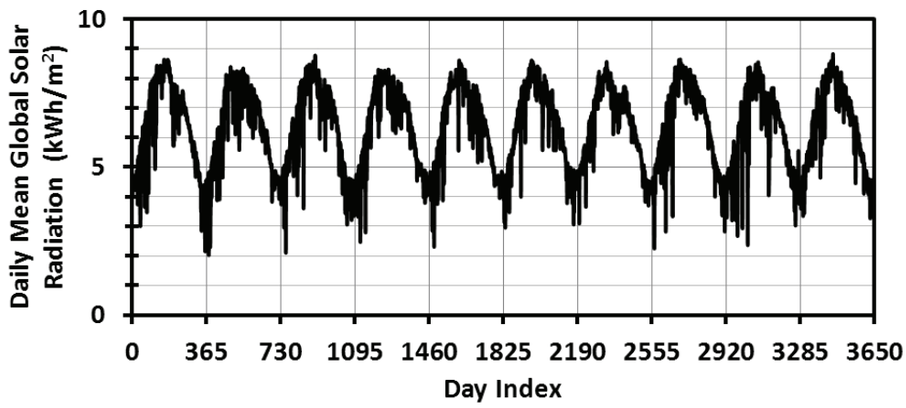


Fig. 4. Daily mean Global Solar Radiation (GSR) in Al-Ain, UAE for years 1995-2004

Coefficient	a_1	a_2	a_3	a_4	a_5
Value	-1.1973	0.0971	0.0513	0.3850	- 0.0052

Table 6. Regression coefficients of eq. (4)

The regression model of eq. (4) yields a deterministic coefficient $R^2 = 0.7824$ and MSE = 0.4247. In MATLAB, the regression between GSR and predictor variables (T, W, SSH, RH) is done using the Statistics Toolbox command

```
>> regcoef= robustfit(xdata, GSR )
```

Where '*xdata*' is an array matrix containing the four predictors, i.e. xdata=[T, WS, SSH, RH], and '*regcoef*' is a vector returning the regression coefficients between GSR and the four predictor weather variables resulting from the robust multi-linear regression process. The correlation between the regression component $GSR_{regression}$ and GSR can be found using

```
>> corr(GSR, GSRregression)
```

Step 5. Trend and Seasonality components of the residual regression error
The residual error from the multivariate linear regression model in eq. (4), denoted by $GSR_{residue1}$ (t) = GSR - $GSR_{regression}$(t), is shown in Fig. 5. The objective is to first examine it for trends and/or seasonality and decompose it to yield:

$$GSR_{residue1}(t) = GSR_{trend}(t) + GSR_{seasonal}(t) + GSR_{residue2}(t) \quad (5)$$

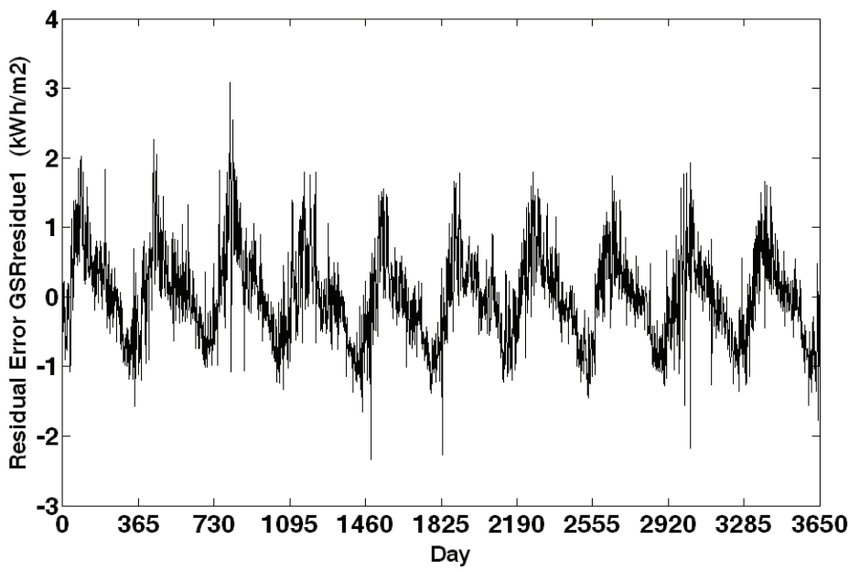


Fig. 5. Residual term of mean daily GSR after subtracting regression model of eq. (4)

The trend component is extracted in MATLAB using the statement

```
>> gsrdet=detrend(gsrresidue1`); % This gives GSRresidue1 with trend removed
>> gsrtrend=gsr-gsrdet;
```

The trend equation can be deduced using the *cftool* command in MATLAB that invokes the curve fitting GUI and then performs a linear polynomial fit to yield the linear trend equation:

$$GSR_{trend}(t) = 0.0771 - 4.2211E-05 * t \quad (6)$$

The mathematical model for the seasonal component, i.e. $GSR_{seasonal}$(t), is found from MATLAB using the FFT algorithm. The resulting model is:

$$GSR_{seasonal}(t) = \sum_{k=1}^{\frac{NP-1}{2}} \left[a_k \cos\left(\frac{2\pi}{NP} kt\right) + b_k \sin\left(\frac{2\pi}{NP} kt\right) \right] \quad (7)$$

The MATLAB statement to generate the Fourier Coefficients for a periodic sequence (period=NP=365) of data x[n], n=1, 2,.., N=3650 is done using:

```
>> N=3650;
>> y=fft(x,N);   % here x is an array representing the de-trendedGSR residual
```

The overall decomposition leads to the following regression model:

$$\mathrm{GSR(t)} = \mathrm{GSR}_{\mathrm{regression}}(\mathrm{t}) + \mathrm{GSR}_{\mathrm{trend}}(\mathrm{t}) + \mathrm{GSR}_{\mathrm{seasonal}}(\mathrm{t}) + \text{residual Error(t)} \quad (8)$$

where the multivariable linear regression, trend and seasonal components are described in eqs. (4), (6) and (7), respectively. The overall residual error of the time-series regression-based model is shown in Fig. 6.

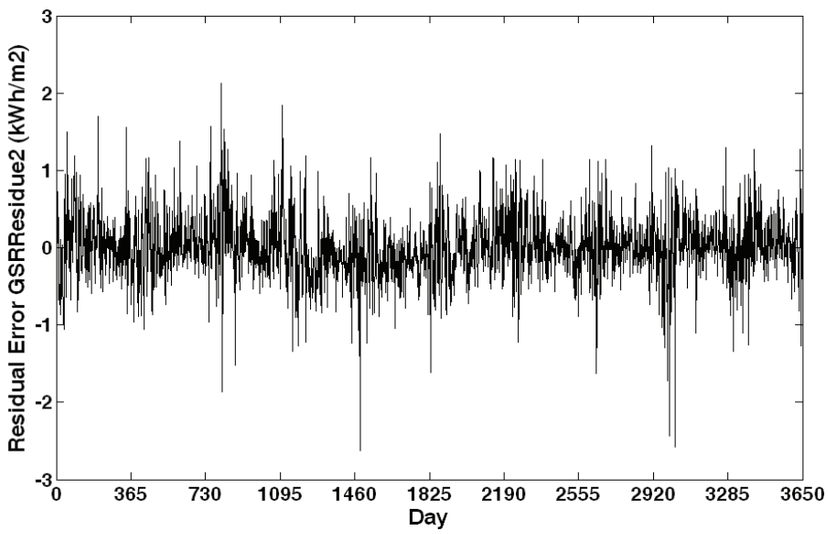


Fig. 6. Residual mean daily global solar radiation error for 1995-2004 data of Al-Ain, UAE

2.1.2.1 Time-series ARMA modelling for the residual term (stochastic component)

The Box-Jenkins models (Box & Jenkins, 1994) are only applicable to stationary time series. The identification of an appropriate box-Jenkins model for a particular time-series would first require a check for stationarity. If the residual term exhibits a normal distribution behavior with zero mean and constant variance then it resembles white noise error and there is no need for further ARIMA modeling.

The behavior of the ACF and PACF plots can help identify the ARMA model that best describes the resulting stationary time-series. Table 7 summarizes the ARMA model selection criteria (Enders 2010). In general, if the ACF of the time series value either cuts off or dies down fairly quick, then the time series values should be considered stationary. On the other hand, if the ACF dies down extremely slow, then the time series values may be considered non-stationary. If the model is adequate then these plots should show all spikes within the 95 % Confidence Interval (CI) bounds ($\pm 1.96 / \sqrt{N}$) where N is the sample size. If the series results non-stationary, one could try to apply differencing or log- transformation and then check if these make the series stationary. A stationary time-series would have a quasi-normal distribution with zero mean and constant variance.

Fig. 7 shows the ACF and PACF plots for the residual component as obtained from Minitab. Note that the ACF decays in an oscillating form after few lags within CI bounds thus implying a fairly stationary time-series. Differencing the residual component made the ACF and PACF more unstable with further deterioration of their behavior thus implying that differencing is inappropriate for this data. The PACF plot cuts off quickly after 1 lag indicating that an AR(2) or higher could be adequate.

Model	ACF	PACF
AR (p)	Spikes decay towards zero. Coefficients may oscillate.	Spikes decay to zero after lag p
MA (q)	Spikes decay to zero after lag q	Spikes decay towards zero. Coefficients may oscillate.
ARMA (p,q)	Spikes decay (either direct or oscillatory) to zero beginning after lag q	Spikes decay (either direct or oscillatory) to zero beginning after lag p

Table 7. Behavior of ACF and PACF for each of the general non-seasonal models

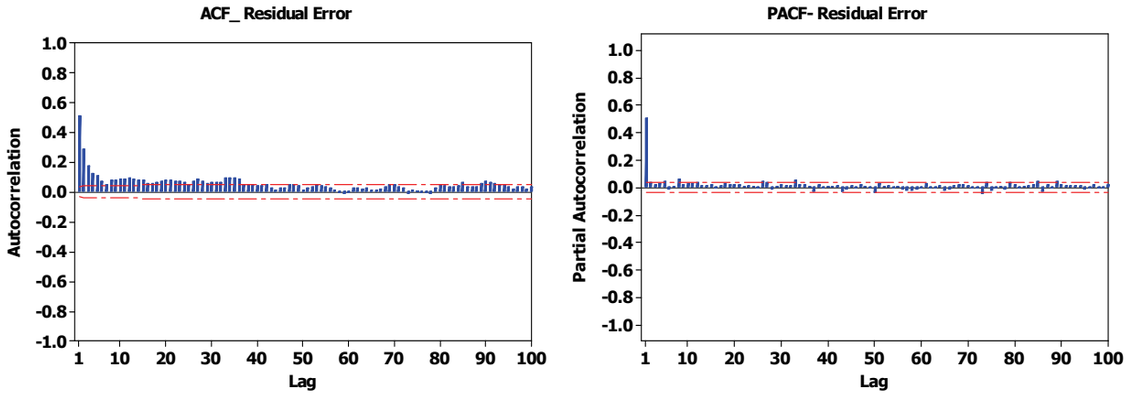


Fig. 7. Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) for residual component of the mean daily GSR data

2.1.2.2 Description of ARMA modeling process

The time-series analysis of the residual stochastic component of the mean daily GSR data is conducted using SPSS and Minitab with help of MATLAB. The nature of the ARMA model used is first described followed by explanation of the model selection process with diagnostic measures used to validate the selected model.

The non-seasonal autoregressive-moving average of order (*p*, *q*), e.g. ARMA(p, q) is described by the equation (Enders, 2010):

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q} \epsilon \quad (9)$$

where δ is a random shock and $\phi_1, \phi_2, \ldots, \phi_p$ are unknown autoregressive model coefficients that are estimated from sample data. The constant parameter is proportional to the mean value μ which is zero in our case. $\theta_1, \theta_2, \ldots, \theta_q$ are unknown moving average model coefficients that depend on the sample data. Moreover, $\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_{t-q}$ are statistically independent random shocks assumed to be randomly selected from a normal distribution with zero mean and constant variance.

The ARMA(p, q) model parameters $(\phi_i, i = 1,2,\ldots,p \text{ and } \theta_j, j = 1,2,\ldots,q)$ described in eq. (9) are estimated using Least Square methods with known values of y_t representing the residual component data. The estimated parameters of the selected ARMA(p, q) models should have t-values higher than 2.0 in order to be judged significantly different from zero at the 5 % level. Moreover, the coefficients should not be strongly correlated with each other in order

to yield a parsimonious ARMA model, whilst passing the diagnostic checks. A parsimonious model is desirable because including irrelevant lags in the model increases the coefficient standard errors and therefore reduces the t-statistics. Models that incorporate large numbers of lags tend not to forecast well as they fit data specific features, explaining much of the noise or random features in the data. Therefore, model coefficients with p-values higher than 0.05 are insignificant and should be eliminated to avoid over-fitting as they have little effect on the prediction model (Enders, 2010). Different models can be obtained for various combinations of AR and MA individually and collectively. The best model is selected using the following diagnostics:

(a) Low Akaike Information Criteria (AIC)/ Schwarz-Bayesian Information Criteria (SBC, BIC)

These model parameters are dependent on the data sample size, model mean-square error (MSE), and the (p, q) values of the ARMA model. Their definition can be found in MATLAB help (MATLAB, 2010) or (Enders, 2010). SBC selects the more parsimonious model and is better than AIC for large samples. *The best model should have the lowest AIC/SBC value and the least MSE*.

(b) Plot of residual autocorrelation function (ACF)

The appropriate ARMA model, once fitted, should have a residual error whose ACF plot varies within the 95% CI bounds ($\pm 1.96 / \sqrt{N}$) where N is the number of observations upon which the model is based.

(c) Non-significance of auto correlations of residuals via Portmanteau tests (Q-tests based on Chi-square statistics) such as Box-Pierce or Ljung-Box tests(White noise tests)

Once the optimal ARMA(p,q) model for the residual GSR time-series is selected, there is a need to check the white noise test if the ACF/PACF correlograms show significant spikes at one or more lags that could be just by chance. These tests indicate whether there is any correlation in the time-series or whether the abnormal spikes encountered in the ACF and PACF of the residual error are just a set of random, identically distributed variables overall. The Ljung-Box Q-statistics can be used to check if the residuals from the ARMA(p, q) model behave as a white-noise process (Ljung & Box, 1978; Enders, 2010). Ljung & Box use the Q statistic:

$$Q = (n)(n+2)\sum_{k=1}^{K} \frac{r_k^2}{n-k} \tag{10}$$

which yields a more accurate variance of ACF [variance becomes (n-K)/n² instead of 1/n] compared to the statistic defined earlier by(Box & Pierce, 1970). K is the degrees of freedom representing the maximum lags considered (normally 20). n = N-d with N being the number of data points and d the degree of differencing (no differencing is assumed in this work so d = 0), and r_k is the sample ACF at lag k.

Under the null-hypothesis that all values of autocorrelation $r_k = 0$, the Q Statistic is compared to critical values from chi-square distribution χ^2 distributed with K-degrees of freedom. If the model is correctly specified, the residuals should be uncorrelated and Q should be small and consequently the probability value should be large. A white noise process would ideally have Q = 0. Therefore, if $Q > \chi^2_{DF,\alpha}$ at the specified DF and significance level α, then the we can reject the null hypothesis.

Several ARMA models were analyzed based on the recommended criteria and two models surfaced out to be the best, namely, ARMA (2,1) and ARMA (4,3). Any further increase in the q-coefficients above 3 (q > 3) lead to over-fitting as witnessed by p-values exceeding 0.05. The best parsimonious model obtained based on the suggested diagnostics is the ARMA(2,1) model. Table 8 shows the ARMA (2, 1) model parameters obtained from SPSS. All the estimated model coefficients have *p* -value less than 0.05 (α= 5 %). This implies that all the coefficients of the selected Box-Jenkins model are significant since the null hypothesis H_0: ϕ= 0 (AR) or θ=0 (MA) can be rejected for the preset significance level α (can be chosen as 0.05 or 0.01). Table 9 shows the model fit statistics for the ARMA (2,1) model. The Ljung-Box statistic value **Q = 15.462** at lag 18 has a corresponding **p = 0.419** > 0.05. Hence, we cannot reject the adequacy of model by setting α = 0.05. The resulting ARMA (2,1) model will therefore yield a residual error that resembles white noise error. The ACF and PACF plots of the residual error of the ARMA(2, 1) model, shown in Fig. 8, vary within the 95% CI bounds. The spikes in the ACF and PACF at lag 7 are due to random events and thus cannot be explained. However, since it is a 95% confidence interval, one can expect this to happen once in every twenty lags and so we will not be concerned with this.

Parameter	Lag	Estimate	Standard Error	t-value	Sig. (p-value)
AR (1)	1	1.4504	0.0189	76.7886	0.000
AR (2)	2	-0.4568	0.0172	-26.5608	0.000
MA (1)	1	0.9699	0.0100	97.0868	0.000

Table 8. SPSS least square estimation of the model parameters for ARMA(2, 1) model of residual GSR component

Model Fit statistics								Ljung-Box Q(18)		
Stationary R^2	R^2	RMSE	MAPE	MAE	MaxAPE	MaxAE	Norm. BIC	Statistics	DF	Sig.
.265	.265	.337	458.224	.238	240313.424	2.491	-2.168	15.462	15	.419

Table 9. ARMA(2, 1) model statistics

Another method to ensure that the selected ARMA model yields a stationary residual error is by checking the **Unit-root rule** for stationarity. Assume that the lag operator in eq. (9) is B. Then $By_t = y_{t-1}$. Eq. (9) can then be written in the form (assume zero mean; δ= 0):

$$\left(1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p\right) y_t = \left(1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q\right)\epsilon_t \tag{11}$$

or

$$\phi(B) y_t = \theta(B)\epsilon_t \tag{12}$$

where

$$\phi(B) = 1-\sum_{i=1}^{p}\phi_i B^i \;; \quad \theta(B) = 1-\sum_{i=1}^{q}\theta_i B^i \tag{13}$$

The stationarity of the time-series sequence y_t requires that all the roots of the AR(p) coefficients polynomial ϕ (B) should lie outside the unit circle (Enders, 2010).

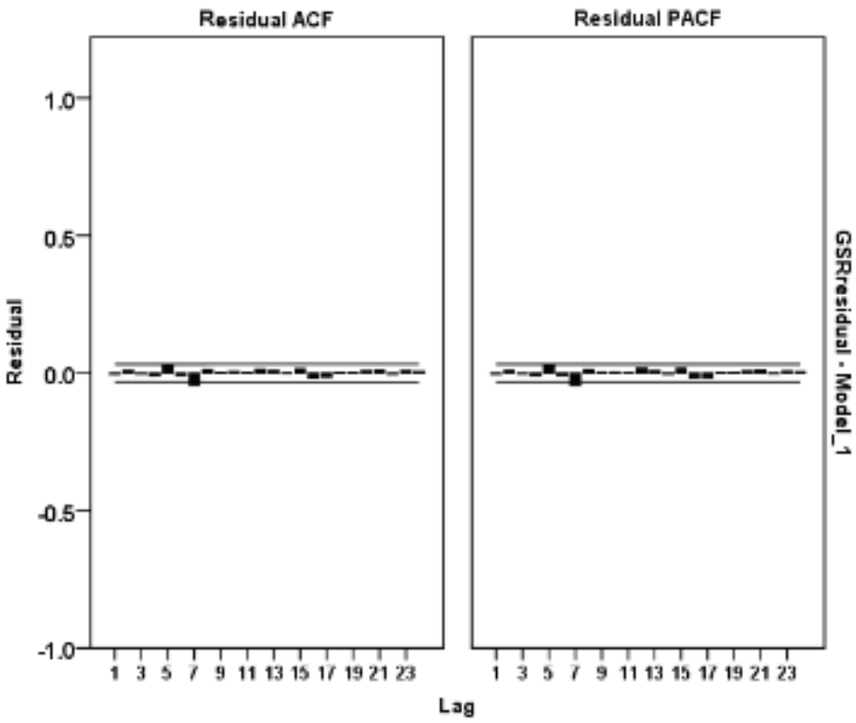


Fig. 8. ACF and PACF plots for the residual error of the ARMA(2, 1) model

The MATLAB function "**roots(c)"** computes the roots of the polynomial P(x) whose coefficients are the elements of the vector **c**. If **c** has (n+1) components, the polynomial is

$$P(x)=c(1)*x^n+ \ldots + c(n)*x + c(n+1).$$

Next, the ARMA(2,1) model parameters $\phi(B)$ obtained from SPSS are used in MATLAB to perform the stationarity unit-root test as follows :

```
>> phi=[1.4504   -0.4568];
>> phiB=[-phi(end:-1:1), 1]
    phiB=
    0.4568  -1.4504   1
>> roots(phiB)   % find roots of polynomial with coefficients phiB
    ans =
       2.1631
       1.0120
```

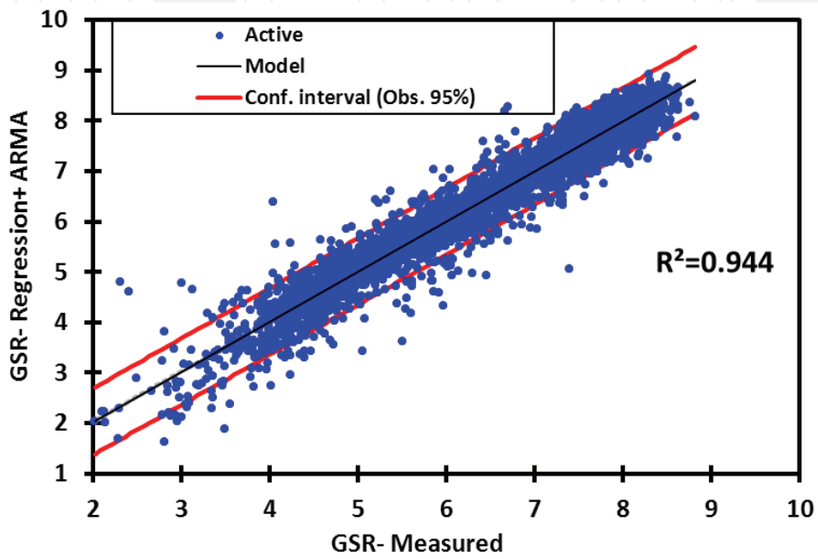
Note that all roots of $\phi(B)$ lie outside the unit circle thus implying a stationary ARMA(2, 1) model. Once the ARMA(p, q) model is selected and checked for stationary residual error, the GSR regression model involving the ARMA model becomes:

$$GSR(t) = GSR_{regression}(t) + GSR_{trend}(t) + GSR_{seasonal}(t) + GSR_{ARIMA}(t) + WN(t) \qquad (14)$$

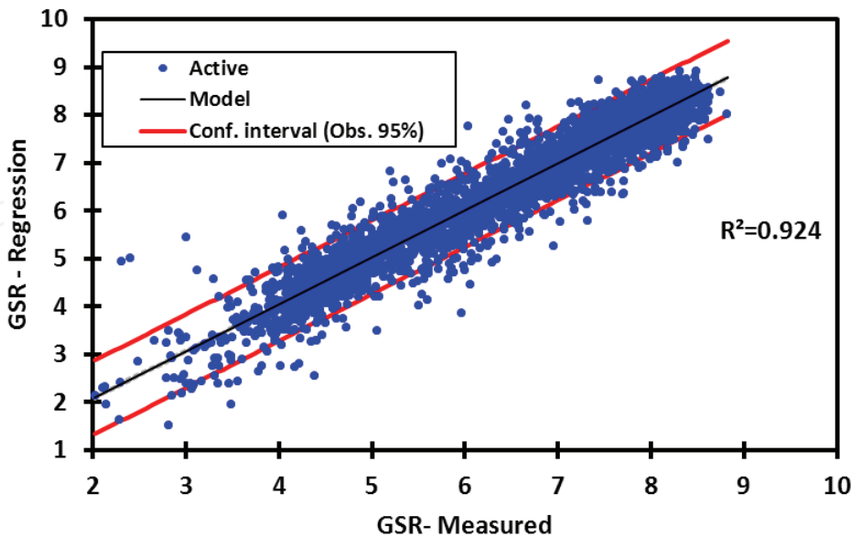
where WN(t) is the final residual error which resembles white noise with zero mean and constant variance.

2.1.2.3 Comparison of Time-series regression model with measured data

Use cftool MATLAB GUI toolbox to study the correlation between the regression model without (eq (8)) and with ARMA modeling (eq (14)) and the measured data set for years 1995-2004. This MATLAB tool sketches the data and finds the polynomial fit and descriptive statistics as depicted in Fig. 9 (a-b). The predicted time-series regression model with and without ARMA modeling yields deterministic coefficients of 94.4% and 92.4 %, respectively, as shown in Fig.9 (a)-(b).The ARMA-based model yields lower error statistics RMSE, MABE and MAPE and hence will do a better forecasting job.



(a)



(b)

Fig. 9. MATLAB "cftool" GUI comparison between regression and measured data for years 1995-2004 (a) with, and (b) without ARMA model

The mean daily GSR data comparison between the measured and the regression model for years 1995-2004 in Al-Ain is shown in Fig. 10. The residual error term of GSR is written as:

$$GSR_{residual}(t) = GSR_{measured}(t) - GSR_{regression}(t) \quad (15)$$

This error is required to have normal distribution with zero mean and constant variance thus resembling a white noise error. The check of normality can be done in SPSS or Minitab or other statistical package. In MATLAB and after invoking the Statistics toolbox, we can use the command 'normplot(x)' where x is the time-series data array. We can also use the Quantile-Quantile Plots (Q-Q Plot) 'qqplot(x,xa)' where x and xa represent the measured and approximate data sequences. A linear relationship between x ad xa suggests that the two samples may come from the same distribution family and hence ensure a normally distributed residual error. Additional statistical studies may be done including ANOVA test and two-sample variance tests with the help of Minitab, SPSS or other statistical software packages. In MATLAB, we can use the Statistics toolbox functions ANOVA1 and ANOVA2.

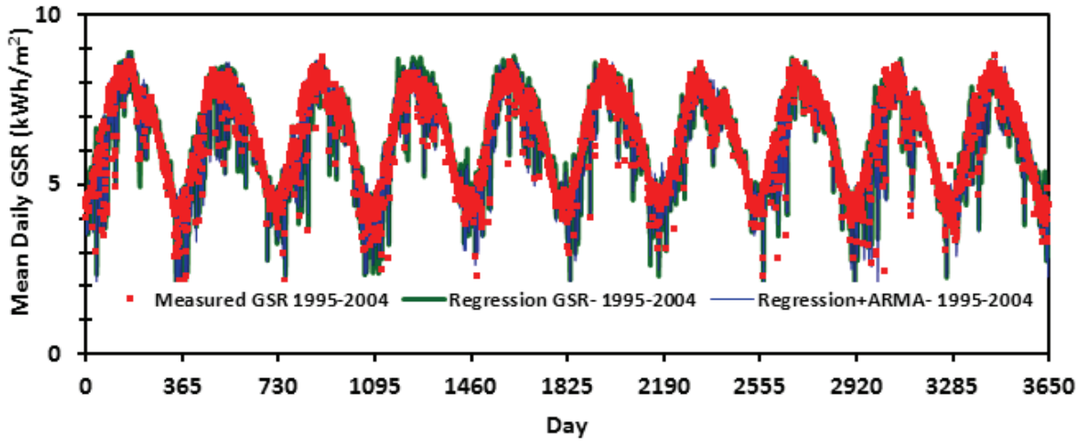


Fig. 10. Mean daily GSR data for years 1995-2004 in Al-Ain, UAE

The residual error of the GSR regression model was checked for normality in the previous section in order to ensure its stationarity for proper prediction of weather data. Minitab or SPSS are used to confirm the normality of the residual error. Fig. 11 shows the normality test results from Minitab showing a residual error with near zero mean (5.8380E-05) and a 0.3364 standard deviation. Note from Fig. 11 that the residual error moves along the normality line between 1 % and 95% percentiles. The computed Skewness (0.204) and Kurtosis (4.85) constants indicate a quasi-normal distribution behavior that is slightly shifted to the left of the mean and with more pointed bell shape than the normal distribution curve.

Table 10 shows the result of running Fisher's two-tailed F-test on the Regression and Measured data for years 1995-2004. This variance ratio test follows the Null hypothesis (H_0) that the ratio between variances is equal to 1 against the alternative (H_a) that the ratio between variances is different from 1. The computed p-value shown in Table 10 is greater than the significance level alpha=0.05 and thus we cannot reject the null hypothesis H_0. The risk to reject the null hypothesis H_0 while it is true is 22.04% and 18.23 % for Regression and Regression with ARMA cases, respectively.

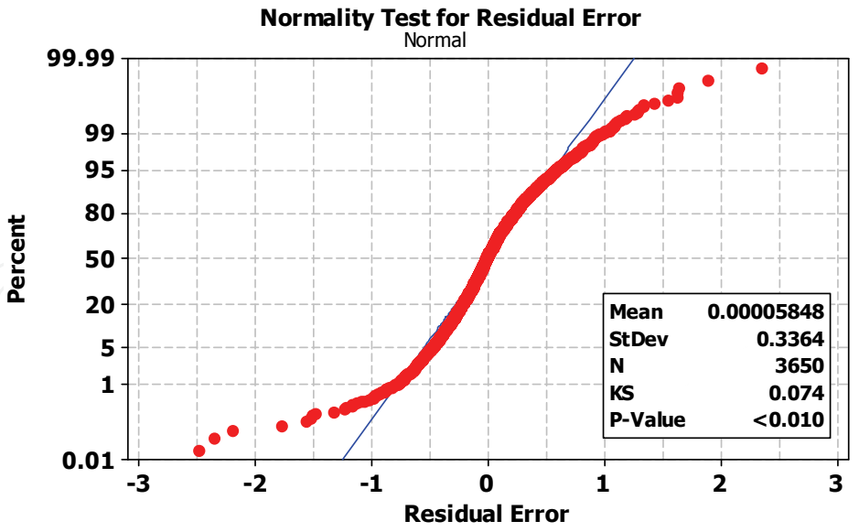


Fig. 11. Normality test for residual error with Minitab

Parameter	Regression vs. Measured	Regression with ARMA vs. Measured
95% CI on ratio of variances	] 0.976, 1.111 [	] 0.979, 1.115 [
Ratio of variances	1.041	1.045
F (Observed value)	1.041	1.045
F (Critical value)	1.067	1.067
DF1, DF2	3649, 3649	3649, 3649
p-value (Two-tailed)	0.2204	0.1823

Table 10. Fisher's two-tailed F-test for the model data of 10 years

Levene's equal variances test can also be applied as shown in Table 11. As the computed p-value in Tables 11 is greater than the significance level alpha=0.05, one cannot reject the null hypothesis H_0. The risk to reject the null hypothesis H_0 in Levene's test while it is true is 55.62% and 43.96% for Regression and Regression with ARMA cases, respectively. Levene's test is mostly used in samples with normal distribution.

Parameter	Regression Vs Measured	Regression with ARMA Vs Measured
F (Observed value)	0.346	0.597
F (Critical value)	3.843	3.843
DF1, DF2	1, 7298	1, 7298
p-value (one-tailed)	**0.5562**	**0.4396**

Table 11. Levene's two-tailed variance test

2.1.2.5 Validation of time-series regression model

In the following section, the time-series regression model is validated with measured test data set for years 2005-2007 in Al-Ain, UAE. A MATLAB code is written to implement the regression model for the input test data (1095 points). The generated regression model test

data is then compared with the test data samples to check the correlation and to study the accuracy of the prediction model. An excellent agreement is noted between the mean daily regression and measured test data for years 2005-2007. The statistical error data computed with MATLAB yields deterministic coefficients R^2 = 92.6% and 90.77% with and without ARMA modeling, respectively, thus indicating good model prediction performance. MATLAB is again used to determine the monthly mean GSR data for the test data period (2005-2007). The resulting monthly mean GSR data comparison between test (measured) and regression model for years 2005-2007 is shown in Fig. 12. Note the excellent agreement between regression model and the test data. Fig. 12 also shows comparison results obtained using Multi-layer Perceptron (MLP) and Radial Basis Function (RBF) Artificial Neural Networks (ANN) obtained by the co-author (Al-Shamisi & Assi, 2011) using the same model and test data sets. Fig. 12 shows a better prediction performance for the regression models over the ANN-based models for Al-Ain test data. The low error parameters (RMSE, MBE, MABE, MBE, and MAPE) obtained for the regression models provide a clear indication of the potential of these techniques for long term GSR data prediction.

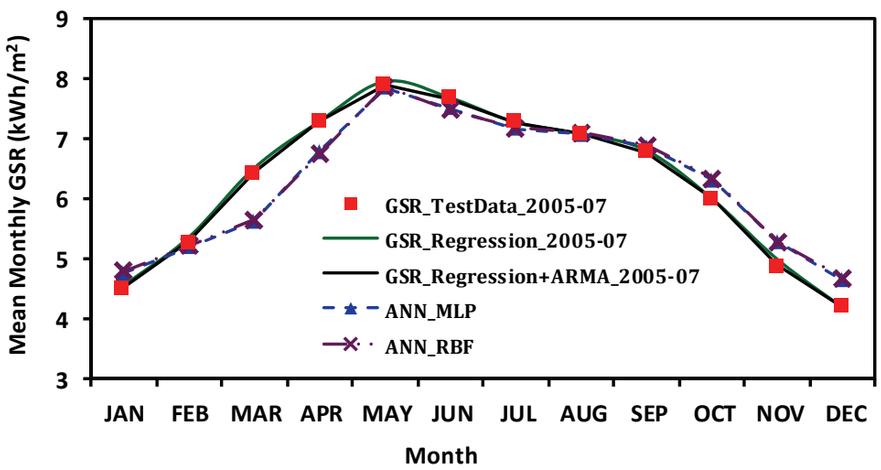


Fig. 12. Monthly mean GSR data comparison for test data 2005-2007 in Al-Ain city

2.2 Definition of the statistical error parameters

This section defines formulas used to compute the statistical error parameters in MATLAB. These parameters attest to the accuracy of the models used for predicting the mean daily global solar radiation (GSR). The error parameters were also computed using EXCEL and SPSS and values agree very well with MATLAB results. The formulas used are:

Mean Average Percentage Error: $\mathrm{MAPE} = 100 \times \frac{1}{N} \sum_{i=1}^{N} \left| \frac{GSR_m^i - GSR_p^i}{GSR_m^i} \right|$

Mean Bias Error: $\mathrm{MBE} = \frac{1}{N} \sum_{i=1}^{N} \left(GSR_m^i - GSR_p^i \right)$

Mean Absolute Bias Error: $\mathrm{MABE} = \frac{1}{N} \sum_{i=1}^{N} \left| GSR_m^i - GSR_p^i \right|$

Root Mean-Square Error: $\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(GSR_m^i - GSR_p^i\right)^2}$

Mean measured value: $\overline{GSR}_m = \frac{1}{N}\sum_{i=1}^{N} GSR_m^i$

Mean predicted value: $\overline{GSR}_p = \frac{1}{N}\sum_{i=1}^{N} GSR_p^i$

Deterministic coefficient: $R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^{N}\left(GSR_m^i - GSR_p^i\right)^2}{\sum_{i=1}^{N}\left(GSR_m^i - \overline{GSR}_m\right)^2}$

where GSR_m^i is the i^{th} measured value, GSR_p^i is the predicted value from the regression model, and N is the total number of data points.

3. Conclusion

This chapter addresses the MATLAB tools employed in finding appropriate prediction models for the mean daily and monthly global solar radiation in the city of Al-Ain, United Arab Emirates. A detailed description of tools used in obtaining the classical empirical regression models as well as time-series ARMA models is presented including computation of error statistics and use of diagnosis tests to validate the selected models. Excellent agreement is observed between the empirical regression and time-series prediction models and measured test data with high deterministic coefficients exceeding 90 % and low MBE, MABE, MAPE and RMSE error statistics that attest to the suitability of these models for long-term weather data prediction. The same MATLAB tools will be used to come up with prediction models for other UAE cities.

4. Acknowledgment

The authors would like to thank the National Center of Meteorology and Seismology (NCMS), Abu Dhabi for providing the weather data. Thanks are also due to Maitha Al-Shamisi for pre-processing the received data. This work is financially supported by UAE University Research Affairs under the contract #1542-07-01-10.

5. References

Abdalla Y.A.G. & Feregh G.M. (1988). Contribution to the Study of Solar Radiation in Abu Dhabi, *Energy Conver. Mgmt.*,Vol. 28, No. 1, pp. 63- 67.

Akinoglu B. G. & Ecevit A. (1990). A further comparison and discussion of sunshine based models to estimate global solar radiation, *Energy*, Vol. 15, pp. 865-872.

Al-Alawi S. & Al-Hinai H. (1998). An ANN-based Approach for Predicting Global Solar Radiation in Locations with no Measurements, *Renewable Energy*, Vol. 14, pp.199–204.

Al Mahdi N., Al Baharna N. S. & Zaki F. F. (1992). Assessment of Solar radiation models for Gulf Arabian Countries, *Renewable Energy*, Vol. 2, No. 1, pp. 65-71.

Ampratwum D. B. & Dorvlo A. S. S. (1999). Estimation of solar radiation from the number of sunshine hours, *Appl. Energy*, Vol. 63, pp. 161-167.

Al-Shamisi M. & Assi A. (January 2011). ANN Global Solar Radiation unpublished results for Al-Ain provided by personal communication.

Assi A. & Al-Shamisi M. (2010). "Prediction of Monthly Average Daily Global Solar Radiation in Al Ain City–UAE Using Artificial Neural Networks", *Proceedings of the 25th European Photovoltaic Solar Energy Conference*, pp. 508-512, Valencia, Spain, September 06-10, 2010.

Assi A., Al-Shamisi M. & Jama M. (2010). Prediction of Monthly Average Daily Global Solar Radiation in Al Ain City–UAE Using Artificial Neural Networks, *Proceedings of the 4th International Conference on Renewable Energy Sources (RES'10),* pp. 109-113, Sousse, Tunisia ,May 03-06, 2010.

Assi A. & Jama M. (2010). Estimating Global Solar Radiation on Horizontal from Sunshine Hours in Abu Dhabi – UAE, *Proceedings of the 4th International Conference on Renewable Energy Sources (RES'10)*, pp. 101-108, Sousse, Tunisia, May 03-06, 2010.

Behrang M.A., Assareh E., Ghanbarzadeh A. & Noghrehabadi A. R. (2010). The potential of different artificial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data, *Solar Energy*, Vol. 84, pp. 1468-1480.

Benghanem M., Mellit A. & Alamri S. N. (2009). ANN-based modeling and estimation of daily global solar radiation data: A case study, *Energy Conversion and Management*, Vol. 50, pp. 1644-1655.

Boccol M., Willington E. & Arias M. (2010). Comparison of Regression and Neural Networks Models to estimate Solar Radiation, *Chilean Journal of Agricultural Research*, Vol. 70, No. 3, pp.428-435, July-Sept., 2010.

Boland J. (1995). Time series analysis of climate variables, *Solar Energy*, Vol. 55, pp. 377–388.

Box G. E. P., Jenkins G. M. & Reinsel G. C. (1994). *Time series analysis: Forecasting and control*, (3rd edition), Prentice-Hall International, Englewood Cliffs, N.J.

Box G. & Pierce D. (1970). Distribution of Autocorrelations in autoregressive Moving Average Time Series Models. *Journal of the American Statistical Association*, Vol. 65, pp. 1509- 1526.

Elagib N. & Mansell M. G. (2000). New approaches for estimating global solar radiation across Sudan, *Energy Conversion Management*, Vol. 41, pp. 419-434.

Elminir H., Areed F. & Elsayed T. (2005). Estimation of solar radiation components incident on Helwan site using neural networks, *Solar Energy*, Vol. 79, pp. 270-279.

Enders W. (2010). *Applied Econometric Time Series* (3rd edition), John Wiley & Sons, ISBN 978-0470-50539-7.

IBM SPSS Statistics for Windows, Version 19.0.0, 2010. SPSS Inc.

Khalil A. & Alnajjar A., (1995), Experimental and Theoretical Investigation of Global and Diffuse Solar Radiation in the United Arab Emirates, *Renewable Energy*, Vol. 6, No. 5-6, pp. 537-543.

Krishnaiah T., Srinivasa Rao S., Madhumurthy K. & Reddy K. S. (2007). A Neural Network Approach for Modelling Global Solar Radiation, *Applied Science Research*, Vol. 3, No. 10, pp.1105-1111.

Ljung G. & Box G. (1978). On a Measure of Lack of Fit in Time Series Models, *Biometrica*, Vol. 65, pp. 297-303.

MATLAB version 7.10.0.499 (R 2010 a), 2010.The MathWorks Inc.

Menges H. O., Ertekin C. & Sonmete M. H. (2006). Evaluation of solar radiation models for Konya, Turkey, *Energy Conversion and Management*, Vol. 47, pp. 3149-3173.

Minitab version 16.1.0, 2010. Minitab, Inc.

Mohandes M., Rehman S., & Halawani T. O. (1998). Estimation of Global Solar Radiation Using Artificial Neural Networks, *Renewable Energy*, Vol. 14, pp. 179-184.

Mohandes M., Balghonaim A., Kassas M., Rehman S. & Halawani T. O. (2000). Use of Radial Basis Functions for Estimating Monthly Mean Daily Solar Radiation, *Solar Energy*, Vol. 68, No. 2, pp. 161-168..

Newland F. J. (1988). A study of solar radiation models for the coastal region of south China, *Solar Energy*, Vol. 31, pp. 227-235.

Podestá G., Núñez G., Villanueva C. & Skanski M. (2004). Estimating daily solar radiation in the Argentine Pampas, *Agricultural and Forest Meteorology,* Vol. 123, pp. 41-53.

Rehman S. & Mohandes M. (2008). Artificial neural network estimation of global solar radiation using air temperature and relative humidity, *Energy Policy*, No. 36, pp. 571-576.

Reikard G. (2009). Predicting solar radiation at high resolutions: A comparison of time series forecasts, *Solar Energy*, Vol. 83, pp. 342–349.

Şahin A. D. (2007). A new formulation for solar irradiation and sunshine duration estimation, *International Journal of Energy Research,* Vol. 31, pp. 109-118.

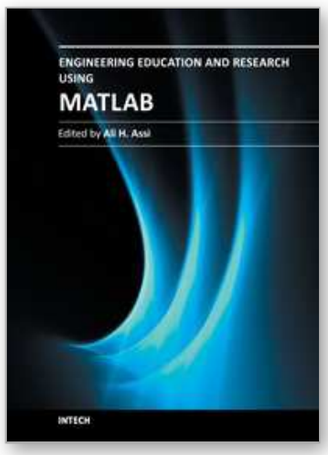
Samuel T. (1991). Estimation of solar radiation for Sri Lanka, *Solar Energy*, Vol. 47, pp. 333–337.

Sulaiman M. Y., Hlaing O. W. M., Wahab M. A. & Sulaiman Z. A. (1997). Analysis of residuals in daily solar radiation time series, *Renewable Energy*, Vol. 11, No. 1, pp. 97-105.

Ulgen K. & Hepbasli A. (2002). Comparison of solar radiation correlations for Izmir, Turkey, *International Journal of Energy Research*, Vol. 26, pp. 413- 430.

Zaharim A., Razali A. M., Gim T. P. & Sopian K. (2009).Time Series Analysis of Solar Radiation data in the Tropics, *European Journal of Scientific Research*, Vol. 25, No. 4, pp. 672- 678.

Zeroual A., Ankrim M. & Wilkinson A. J. (1995). Stochastic modelling of daily global solar radiation measured in Marrakesh, Morocco, *Renewable Energy*, Vol. 6, No. 7, pp. 787-793.



Engineering Education and Research Using MATLAB

Edited by Dr. Ali Assi

ISBN 978-953-307-656-0
Hard cover, 480 pages
Publisher InTech
Published online 10, October, 2011
Published in print edition October, 2011

MATLAB is a software package used primarily in the field of engineering for signal processing, numerical data analysis, modeling, programming, simulation, and computer graphic visualization. In the last few years, it has become widely accepted as an efficient tool, and, therefore, its use has significantly increased in scientific communities and academic institutions. This book consists of 20 chapters presenting research works using MATLAB tools. Chapters include techniques for programming and developing Graphical User Interfaces (GUIs), dynamic systems, electric machines, signal and image processing, power electronics, mixed signal circuits, genetic programming, digital watermarking, control systems, time-series regression modeling, and artificial neural networks.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Hassan A. N. Hejase and Ali H. Assi (2011). MATLAB-Assisted Regression Modeling of Mean Daily Global Solar Radiation in Al-Ain, UAE, Engineering Education and Research Using MATLAB, Dr. Ali Assi (Ed.), ISBN: 978-953-307-656-0, InTech, Available from: http://www.intechopen.com/books/engineering-education-and-research-using-matlab/matlab-assisted-regression-modeling-of-mean-daily-global-solar-radiation-in-al-ain-uae

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen