Available online at www.sciencedirect.com

SciVerse ScienceDirect

Energy Procedia 18 (2012) 1402 – 1412

Energy Procedia

Linear Regression Model in Estimating Solar Radiation in Perlis

S. Ibrahim[b], I.Daut[a], Y.M. Irwan [a*], M. Irwanto[a], N. Gomesh[a], Z. Farhana[a]

[a] *Cluster Power Electronic and Machine Design, School of Electrical System Engineering, University Malaysia Perlis (UniMAP), Malaysia,*
[b] *Institute of Mathematical Science, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia.*

Abstract

Statistical models for predicting the solar radiation have been developed. In any prediction of the solar radiation, an understanding of its characteristics is of fundamental importance. This study presents an investigation of a relationship between solar radiation and temperature in Perlis, Northern Malaysia for the year of 2006. To achieve this, the data are presented in daily averaged maximum and minimum air temperature, and daily averaged solar radiation. Since the scatter plots represent the straight line, the linear regression model was selected to estimate the solar radiation. It was found that the linear correlation coefficient value is 0.7473 shows that a strong linear relationship between solar radiation and temperature. The analysis of variance R^2 is 0.5585 that is; about 56 percent of the variability in temperature is accounted for by the straight-line fit to solar radiation. Based on the results, the fitted model is adequate to represent the estimation of solar radiation.

Keywords: *Solar radiation; Temperature; Linear regression model; Statistical analysis;*

1. Introduction

Energy is one of the essential inputs for economic development and industrialization. Fossil fuels are the main resources and play a crucial role to supply world energy demand. However, fossil fuel reserves are limited and usage of fossil fuel sources has negative environment impact. Therefore, management of energy sources, rational utilization of energy, and renewable energy source usage are vital [1][2]. Renewable energy has an increasing role in achieving the goals of sustainable development, energy security and environmental protection. Nowadays, it has been recognized as one of the most promising

* Corresponding author. Tel.:+60104004700; fax: +604-9798903.
E-mail address: irwanyusoff@unimap.edu.my

 doi:10.1016/j.egypro.2012.05.156

clean energy over the world because of its falling cost, while other renewable energy technologies are becoming more expensive [3].

1.1 Solar radiation

Solar radiation is the result of fusion of atoms inside the sun. Part of the energy from the fusion process heats the chromosphere, the outer layer of the sun that is much cooler than the interior of the sun, and the radiation from the chromosphere becomes the solar radiation incident on the earth [4]. Wind energy is produced by continuously blowing wind and can be captured using wind turbines that convert kinetic energy from wind into mechanical energy and then into electrical energy [5].
When the solar radiation enters the earth's atmosphere (Fig. 1), a part of the incident energy is removed by scattering or absorption by air molecules, clouds and particulate matter usually referred to as aerosols. The radiation that is not reflected or scattered and reaches the surface directly in line from the PV module is called beam radiation. The scattered radiation which reaches the ground is called diffuse radiation. Some of the radiation may reach a receiver after reflection from the ground, and is called the albedo. The total solar radiation on a horizontal surface of PV module consisting three components is called global irradiance. When the skies are clear and the sun is directly in line from the PV module, the global irradiance is about 1000 W/m2 [6]. Although the global irradiance on the surface of the earth can be as high as 1000 W/m2, the available radiation is usually considerably lower than this maximum value due to the rotation on the earth and climate condition (cloud cover), as well as by the general composition of the atmosphere. For this reason, the solar radiation data is the most important component to estimate output of photovoltaic systems [4] [7] [8]. Solar radiation is greater than 3 kWh/m2 indicates that the sky is clear, its intensity very high and very good for PV application [9].

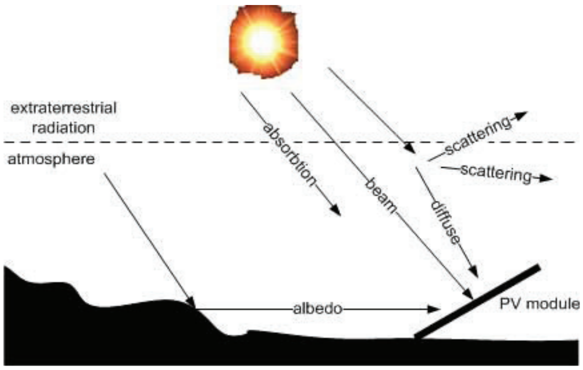

Fig. 1 Solar radiation in the earth's atmosphere

1.2 Simple Linear Regression model

Regression analysis is a statistical technique for investigating and modelling the relationship between variables [11]. In fact, the regression analysis is the most widely used statistical technique. The simple linear regression model used is a model with a single independent variable *x* that has a relationship with a response variable *y* that is a straight line. This simple linear regression model is given by

$$y = \beta_0 + \beta_1 x + \varepsilon$$

where the intercept β_0 and the slope β_1 are unknown constant and ε is a random error. The errors are assumed to have mean zero and unknown variance σ^2. The parameters β_0 and β_1 are unknown and must be estimated using sample data. The simple linear regression equation is also called the *least squares* regression equation. It tells the criterion used to select the best fitting line, namely the sum of the *squares* of the residuals should be *least*. That is, the least squares regression equation is the line for which the sum of squared residuals $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ is a minimum.

2. Data and Methods

2.1 Latitude and climate of Perlis, Northern Malaysia

Based on Malaysia Meteorological Department [10], Malaysia naturally has abundant sunshine and thus solar radiation. However, it is extremely rare to have a full day with completely clear sky even in periods of severe drought. The cloud cover cuts off a substantial amount of sunshine and thus solar radiation. On the average, Malaysia receives about 6 hours of sunshine per day. Based on Meteorological Station in Chuping, Perlis (60 29' N , 1000 16' E) as shown in Fig.2 has about 795 square kilometers land area, 0.24% of the total land area of Malaysia, with a population about 204450 people [10]. Perlis's climate is tropical monsoon. Its temperature is relatively uniform within the range of 21°C to 32°C throughout the year. During the months of January to April, the weather is generally dry and warm. Humidity is consistently high on the lowlands ranging 82% to 86% per annum. The average rainfall per year is 2,032 mm to 2,540 mm and the wettest months are from May to December. In this research, the data are presented in daily averaged maximum and minimum temperature, and daily averaged solar radiation.

Fig. 2 Map of Perlis has latitude 60° 29' N

2.2 Least Squares Method

Suppose that for any observation, $(x_1, y_1),(x_2, y_2),\ldots,(x_n, y_n)$ be a pair of random variables. To predict *y*, the parameters β_0 and β_1 must be estimated, so that the sum of the square of the differences between the observations y_i and the straight line is minimum. It gets this name because the resulting values of β_0 and β_1 minimize the expression $\sum_{i=1}^{n}\left(y_i - [B_0 + B_1 x_i]\right)^2$. In formulas, the interpolating straight line may be written as:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad i=1, 2, ..., n \tag{1}$$

and the coefficients that minimize the square of the distance between the line end the points are given by:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{2}$$

$$\beta_1 = \frac{\sum_{i=1}^{n} y_i\, x_i - \frac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n}}{\sum_{i=1}^{n} x_i^{\,2} - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}$$

where

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad \text{and} \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

are the averages of y_i and x_i, respectively. Therefore, β_0 and β_1 are the least squares estimators of the intercept and slope. The residuals ε are the differences between the observed and the predicted values $y_i - \hat{y}_i$, *i*=1, 2, ..., *n*.

The fitted simple linear regression model is given by

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_i \tag{3}$$

The correlation coefficient *r* evaluates the goodness of the fitting of data considered and the standard error measures, *s* is calculated. The correlation coefficient value can vary in the range –1 and +1, for the strong correlation between the two variables *x* and *y*. If the value is zero there is not any linear correlation between the two variables. The calculation of *r* and *s* are respectively as follow

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]\left[n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right]}} \tag{4}$$

$$s = \frac{\sum_{i=1}^{n} y_i^2 - B_0\sum_{i=1}^{n} y_i - B_1\sum_{i=1}^{n} x_i y_i}{n} \tag{5}$$

The coefficient of determination, R^2 is given by

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \overline{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{6}$$

where $0 \leq R^2 \leq 1$. R^2 is often called the proportion of variation explained by the regressor *x*. Values of R^2 that are closed to 1 imply that most of the variability in *y* is explained by the regression model [11].

3. Result and Discussion

3.1 Data Analysis and Parameter estimation

Based on the average between maximum and minimum air temperature in Perlis, the solar radiation for the year of 2007 can be estimated using linear regression model. The daily averaged maximum and minimum air temperature and daily averaged solar radiation throughout the year of 2007 in Perlis are shown in Fig. 3 and 4, respectively.

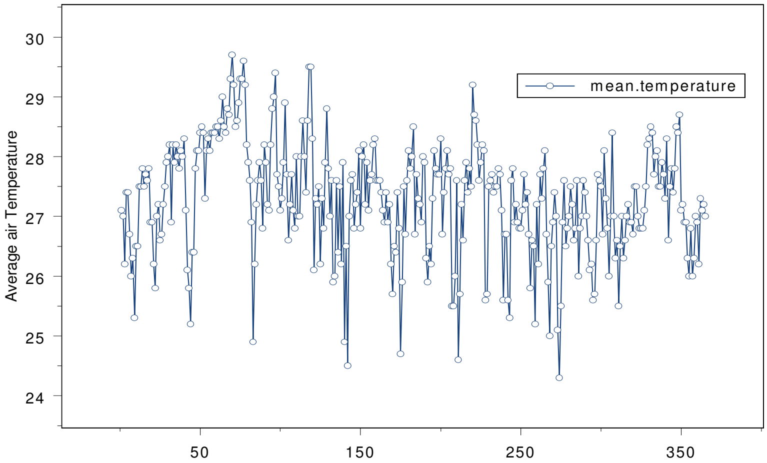

Fig. 3 The graph of average air temperature

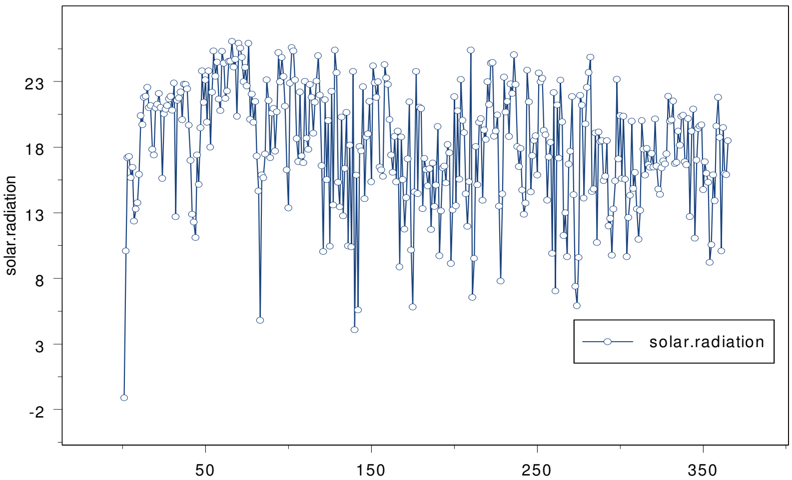

Fig. 4 The graph of solar radiation

After that, the analysis part is to obtain a simple relation between solar radiation (y) and air temperature (x). As a first step in assessing whether or not there appears to be a strong relationship between these two variables, we make a scatter plot as shown in Fig. 5. The scatter plot consists of the data points, where there is a positive and roughly linear relationship between solar radiation and the increase in air temperature. The three exceptional data points (observations no. 1, 22 and 198) are well detached from the remainder of the data. These data points are called outliers.

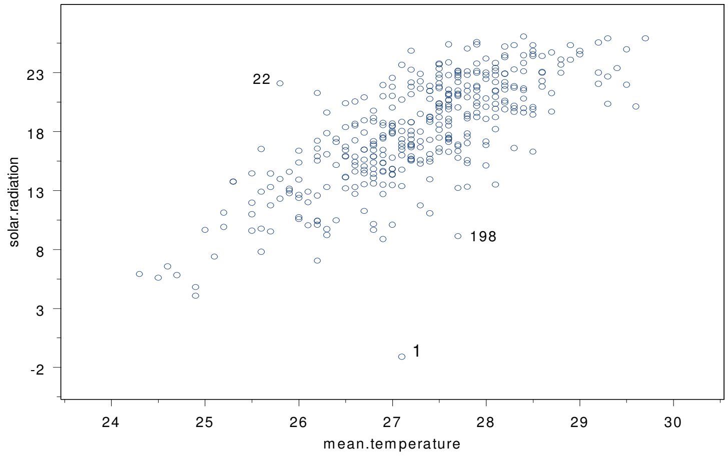

Fig. 5 The scatter plot of solar radiation versus air temperature

A simple linear regression model is assumed, and the estimation of parameters in the regression model is calculated using least squares method as given in eq. (2). The value of $\hat{\beta}_0$ and $\hat{\beta}_1$ is -80.4560 and 3.6100 respectively. The least squares fit to the solar radiation data is

$$\hat{y} = -80.4560 + 3.6100x$$

where $\hat{y}$ is the estimated value of solar radiation corresponding to the air temperature of *x* cases. The fitted equation is plotted in Fig. 6.

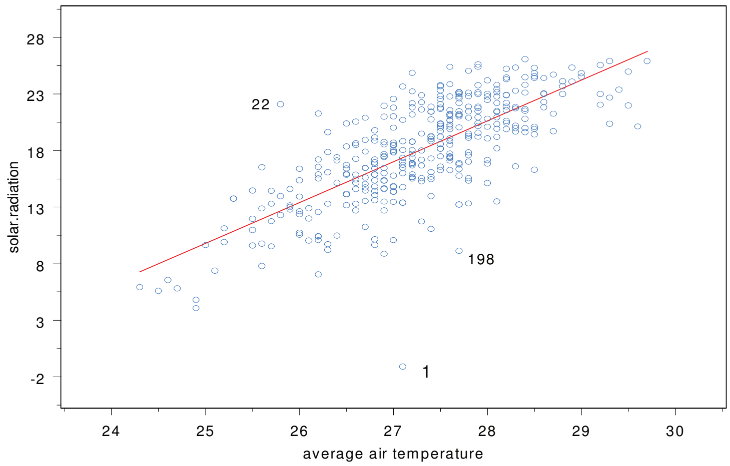

Fig. 6 The scatter plot with the fitted line of air temperature versus solar radiation

The linear correlation coefficient value in this case is 0.7473. It is shows that a strong linear relationship between solar radiation and air temperature. This is because the solar radiation and air temperature is directly proportional. The analysis of variance for this model $R^2 = 0.5585$; that is, about 56 percent of the variability in temperature is accounted for by the straight-line fit to solar radiation.

3.2 *Measure of Model Adequacy*

To diagnostic and checking the adequacy of the model, the assumption of the regression analysis are studied. The residual analysis have defined as

$$e_i = y_i - \hat{y}_i, \; i = 1, 2, \dots n$$

where y_i is an observation and $\hat{y}_i$ is the corresponding predicted values. Analysis of residuals is an effective method for discovering several types of model deficiencies. The normal probability plot represented that the errors (residuals) are normally distributed and shown in Fig. 7 since the points lie approximately along the straight line. Here, a common defect that shows up the occurrence of three large residuals, indicated that the corresponding observations as outliers [11] [12].

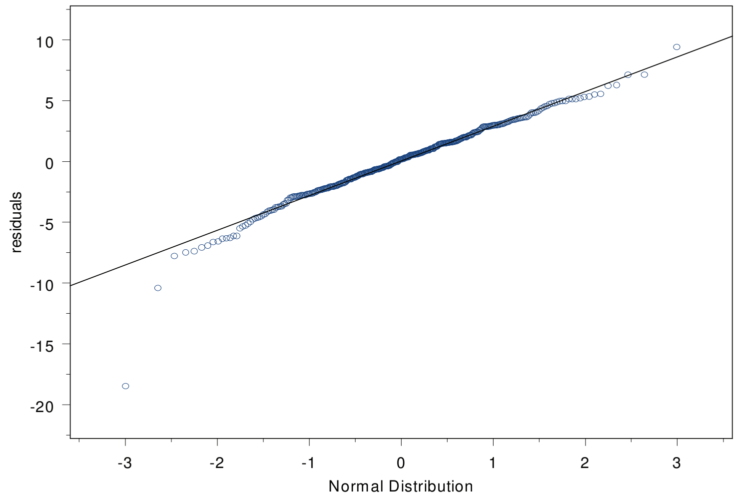

Fig. 7 The graph of the normal probability plot of residuals

A plot of residuals e_i versus the corresponding fitted values $\hat{y}_i$ is useful for detecting several common types of the model inadequacies. Plot of e_i versus $\hat{y}_i$ is shown in Fig. 8. Based on this figure, the patterns show the variance of the errors is constant and stable. Therefore, the residual plots are satisfactory and the model is adequate.

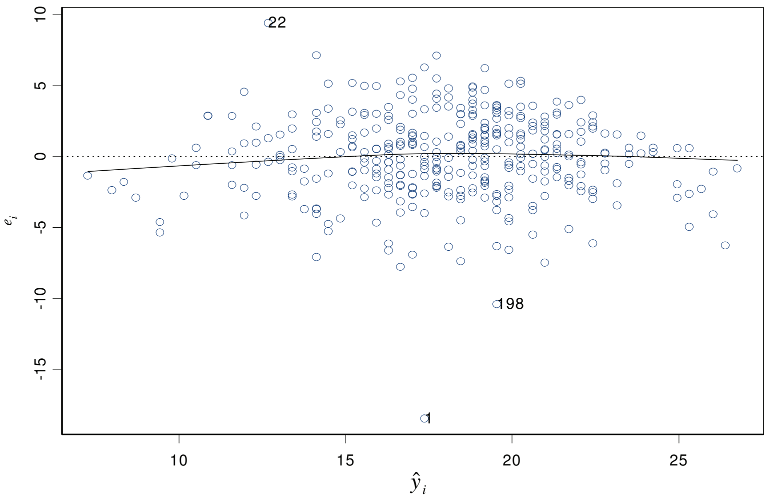

Fig. 8 The graph of the residuals plot

There are three observations seems to be an outliers. The box plot graph in Fig. 9 also represents that three outliers appears the outer of the boundaries.

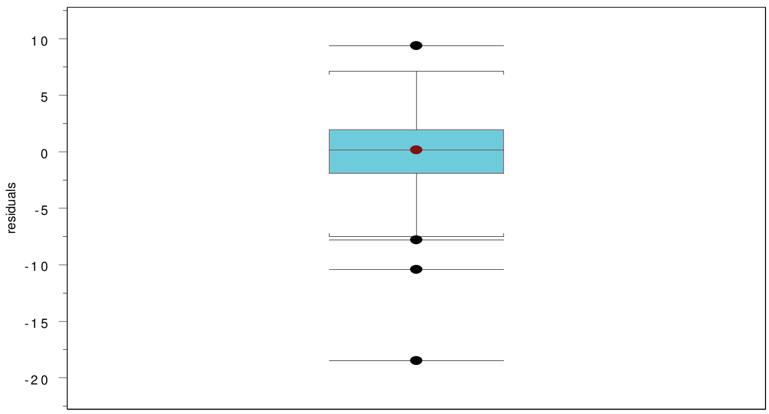

Fig. 9 The graph of the box plot

The outlier will be affecting the estimated coefficients, fitted values, residuals and covariance matrix of linear regression models [12] [14]. Therefore, we deleted these observations and fit it again. The fitted model is obtained as follow

$$\hat{y} = \text{-}81.6415 + 3.6554x \,.$$

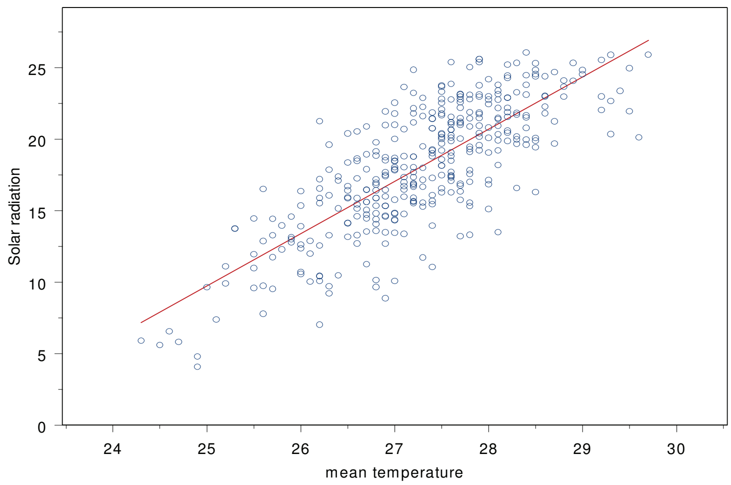

Fig. 10 The graph of the fitted line in the regression model

The Fig. 10 shows that the linear correlation coefficient value between air temperature and solar radiation is 0.7780. It is shows that a strong linear relationship between solar radiation and air temperature. This is because the solar radiation and air temperature is directly proportional. The analysis of variance for this model $R^2 = 0.6054$; that is, about 61 percent of the variability in air temperature is accounted for by the straight-line fit to solar radiation. These results show that the model without outliers is the best fitted model compares to the earliest fitted data contained outliers.

Conclusion

According to results above, the linear regression model can be used to estimate the solar radiation in Perlis, Northern Malaysia. The relationship between average air temperature and solar radiation is linear. The linear correlation coefficient value between air temperature and solar radiation is 0.7780 for without outliers, and 0.7473 with the outliers. These results indicate that the correlation coefficient and the value of R^2 are higher when removing the outliers.

Acknowledgements

The authors wish to thank School of Electrical System Engineering, University of Malaysia Perlis (UniMAP) for the technical and Fundamental Research Grant Scheme 2011 for financial support as well.

References

[1] I. Daut, M. Irwanto , Y.M. Irwan, N. Gomesh, N.S. Ahmad "Combination of Hargreaves method and linear regression as a new method to estimate solar radiation in Perlis, Northern Malaysia" Journal of Solar Energy 85 (2011) 2871–2880.
[2] U. Aynur, B. Figen, "Assesment of wind power potential for turbine installation in coastal areas of Turkey", Renewable and Sustainable Energy Reviews, vol.14:, pp. 1901- 1912, 2010.
[3] O.Ahmed, D. Hanane, S.Roberto, M. Abdelaziz, "Monthly and seasonal of wind energy characteristics at four monitored locations in Liguria region (Italy)", Renewable and Sustainable Energy Reviews, vol. 14, pp. 1959- 1968, 2010.
[4] T. Markvart, "Solar Electricity" John Wiley & Sons, LTD., New York, 1994.
[5] S.A. Ahmed, "Wind energy as a potential generation source at Ras Benas, Egypt", Renewable and Sustainable Energy Reviews, vol.14:, pp. 2167- 2173. 2010.
[6] F. Jiang, "Investigation of Solar Energy for Photovoltaic Application in Singapore", the 8th International Power Engineering Conference (IPEC 2007), pp.86 – 89, 2007.
[7] A. Itagaki, H. Okamura, M. Yamada, "Preparation of Meteorological Data Set Throughout Japan For Suitable Design of PV Systems", 3rd World Conference on Photovoltaic Energy Conversion,, pp. 2074 – 2077, 2003.
[8] A. Mellit, S.A. Kalogirou, S. Shaari, H. Salhi, A.H. Arab, " Methodology for Predictiong Sequences of Mean Monthly Clearness Index and Daily Solar radiation Data in Remote Areas: Application for Sizing a Standalone PV System", Renewable Energy, Science Direct, pp. 1570 – 1590, 2007.
[9] Y. Shijun, Y. Hongxing, "The Potential Electricity Generating Capacity of BIP In Hong Kong", IEEE Xploree, pp. 1345 –1348.
[10] M. Yorukoglu, and A.N. Celik, "A Critical Review on the Estimation of Daily Global Solar Radiation from Sunshine Duration", Energy Conversion & Management, 2006. pp. 2441-2450.
[11] Montgomery, D. C. and Peck, E. A "Introduction to linear regression analysis" 2nd ed. Wiley. New York, 1992.
[12] Barnett, V. and Lewis, T, "Outliers in Statistical Data" Wiley and son, New York, 1994.
[13] Chatterjee. S and Hadi. A. S. "Influential Observations, High Leverage Points, and Outliers in Linear Regression" Statistical Science, Vol. 1, No. 3, pp. 379-393, 1986.
[14] Belsley, D. A., Kuh, E. and Welsch, R. E. "Regression Diagnostics: Identifying influential data and sources of collinearity" John Wiley & Sons, New York, 1980.