Stochastic Model for PV Sensor Array Data

Faris Alfaris
Electrical and Computer Engineering
North Carolina State University
Raleigh, NC 27607
fealfari@ncsu.edu

Ahmad Alzahrani and Jonathan W. Kimball
Electrical and Computer Engineering
Missouri University of Science and Technology
Rolla, MO 65401
kimballjw@mst.edu

Abstract—**Recently, a number of researchers have investigated photovoltaic (PV) system modeling. Modelling a PV panel and its incident solar radiation to predict future trends improves a system's performance. This paper presents a fast, practical method that can be used to predict PV output power. By using present data of weather condition and present output power of the PV system, this predictor is modeled using linear regression analysis. The data from multiple sensors is collected only once before it is correlated to one sensor so that, in the future, only one sensor is needed to collect the data. Several experiments conducted under different weather conditions and different windows sizes of linear regression were completed to validate this method. These results were compared to the Meinel and Meinel model. This method yielded promising results, as the root mean square errors were low.**

I. INTRODUCTION

Tremendous improvements in PV systems have been made in the past few years. These improvements include those made at the operation level. Research into predicting solar energy has become more active as a result of this progress. The ability to predict solar energy increases the economic benefits of such technology that can assist and reduce renewable energy expansion constraints [1]. Researchers need models that can assist in these predictions, thereby improving performance. Both physical and stochastic models can be used to predict sun irradiance. A number of physical models can predict solar irradiance. In this study, the Meinel and Meinel model was used as a reference for the experiment output.

A. Meinel and Meinel Model

The Meinel and Meinel model is the best fit to the sun radiation based on its mean bias error (MBE) and root mean square error (RMSE) tests [2]. This model sets the sunrise and sunset times without using extra equations (as other models do). Declination (δ) is defined first; it is given for any day of the year:

$$\delta = 23.45^{o} \sin\left[\frac{360\ (n-80)}{365}\right] \tag{1}$$

where n represents the day of the year, and 23.45° represents the angle of the earth's polar axis of the earth with the sun. Angles north of the equator are considered positive in the reference; they are negative south of the equator [3]. Figure 1 illustrates the declination at different times of the year.

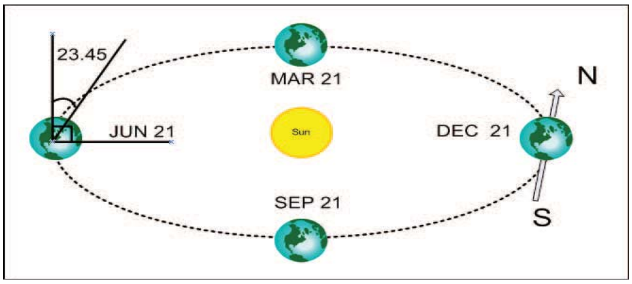

Fig. 1. Declination at different times

Solar noon time can be calculated as

$$t = 12 + \frac{\varphi_L - \phi_n}{15} \times 60 \tag{2}$$

where ϕ_L is the longitude at which the point of interest lies, and ϕ_n is the longitude at which the solar noon occurs (relative to the local time zone). The term (60 min/15°) is used because the earth needs 60 minutes to rotate 15 degrees. An important relationship between the latitude (ϕ), the hour angle (ω), and δ is used to identify the sun's position in terms of α at a given location, date, and time. This relationship is

$$\sin\alpha = \sin\delta\sin\phi + \cos\delta\cos\phi\cos\omega \tag{3}$$

To find the final form of the sun radiation model, the air mass coefficient AM is needed to include the length of the path through the atmosphere. This path length is generally compared to a vertical path that is located directly at sea level. Therefore, the air mass was calculated in [2] as

$$AM = 1/\sin\alpha \tag{4}$$

The energy received from the sun's radiation, at a particular location, on a specific day and time, can be estimated as

$$I = 1367 \times (0.7)^{AM} \tag{5}$$

where AM =1. However, according to Meinel and Meinel [1], a better fit to the observed data is given by

This work was supported, in part, by the National Science Foundation under award ECCS-0900940

$$I = 1367 \times (0.7)^{AM^{0.678}} \quad (6)$$

B. Linear Regression

Regression analysis is a statistical method that is used to form a relationship between two variables. The dependent variable in linear regression methods is modelled as a linear function of independent variables so that the model provides the best fit for the actual data. This method is also frequently used for a time series variable to estimate the next value if the relationship between two sets of data (x and y) is linear. Hence, this method can be applied to sun radiation data, with the sun radiation model (Meinel and Meinel model) used as a reference. Parameters from the regression model can be found through various techniques. Each technique produces different parameter values and, therefore, different estimations. The least squares method is most often used to identify these parameters [4]. A simple linear regression is used here. The system that needs to be modelled has one dependent variable (sun radiation) and one independent variable (time). This model can be written as

$$y_{(k+1)} = \beta_{o,k} + \beta_{1,k} \times x_{k+1} + \varepsilon \quad (7)$$

where β_o is the y- intercept, β_1 is the slope, y is the dependent variable, x is the independent variable, and ε is a random error. These parameters are estimated according to pre-defined criterion. The principle of the least square method is to find the values of β_o and β_1 such that the sum of the squared distance between the actual data and the fitted data reaches the minimum among all possible choices of regression coefficients β_o and β_1 [3][4]. Therefore, depending on the least squares algorithm, the slope and the y- intercept parameters are calculated as

$$\beta_1 = \frac{n\sum_i^k (x_i y_i) - \sum_i^k x_i \sum_i^k y_i}{n\sum_i^k (x_i^2) - \left(\sum_i^k x_i\right)^2} \quad (8)$$

$$\beta_o = \frac{\sum_i^k y_i - \beta_1 \sum_i^k x_i}{n} \quad (9)$$

where $i = k - n$, and k is the number of the current value. For example, if the 25th reading is reached, then k = 25. The coefficient n (the "window size") is the number of previous data that is used to calculate the next value. Different window sizes affect the estimator model's accuracy. However, the window size in the linear regression model presented here, was equal to 3 because its fit accuracy depended on the Mean Squared Error (MSE). Different window size sets are discussed further in the following section. The correlation coefficient is needed for further work. It can be calculated as

$$r = \frac{n\sum(xy) - \sum x \sum y}{\sqrt{\left[n\sum(x^2) - \left(\sum x\right)^2\right]\left[n\sum(y^2) - \left(\sum y\right)^2\right]}} \quad (10)$$

The regression residual plays an important role in the regression method. This residual is defined as the difference between the actual response (y_i) and the fitted response ($\widehat{y_i}$), as presented in (11).

$$e_i = y_i - \hat{y}_i \quad (11)$$

It should be noted that the regression residual was observable, but the error term in the regression model (ε) was unobservable. Therefore, the linear regression method was applied to the difference between the actual sun radiation data and the sun spectrum model (Meinel and Meinel model). MATLAB software was used to generate the random error (ε).

II. Correlation Method and Array data Construction

Another primary goal of this study was to use one sensor's reading to build up the data within the entire array. The slope between both each neighbor's sensors as well as between the correlation coefficients needed to be identified before relationships could be built between each of the sensors. Standard deviation values were used to find the correlation coefficients. Thus, the correlation coefficient between sensor1 (x) and sensor2 (y) was

$$R_{(x,y)} = \frac{\mathrm{cov}(x,y)}{\sigma_x \sigma_y} \quad (12)$$

where $R_{(x,y)}$ is the correlation coefficient between x and y, cov(x,y) is the covariance matrix, and σ_x and σ_y are the standard deviation of x and y, respectively. These two values (slope and correlation coefficient) were used to build all of the array radiation data from only one sensor.

$$i_y = \left[\frac{\sum_{x=j}^{n}\left[i_x \times (R_{(y_x)} + slope_{(y_x)})\right]}{\sum_{x=j}^{n}(R_{(y_x)} + slope_{(y_x)})}\right] + \varepsilon \quad (13)$$

where i_y is the calculated short circuit current, and i_x is the current related to i_y by slope (y_x). The correlation coefficient is $R_{(x,y)}$, where x represents the name of each sensor located side-by-side with i_y. The flow chart in Fig. 2 illustrates the process used to construct the array's data. The construction began at B4 sensor's reading to generate data from sensors B5, B3, A4, and C4. These five sets of data were used to calculate the remaining four sensor's data (A3, C3, A5, and C5). Row 5 (A5, B5, and C5) was used to calculate B6, and row 3 (A3, B3, and C3) was used to calculate B2. The calculation process was the repeated over the entire array. Thus, A6 was calculated from B5, A5, and B6, C6 was calculated from B5, C5, and B6. A2 was calculated from B3, A3, and B2, and C2 was calculated from B3, C3, and B2. Again, rows 6 and 2 were used to calculate B7, and B1, respectively. The same sequence was applied again, so that A7 was calculated from B6, A6, and B7, C7 was calculated from B6, C6, and B7, A1 was calculated from B2, A2, and B1, C1 was calculated from B2, C2, and B1, A8 was calculated from B7, A7, and B8, and C8 was calculated from B7, C7, and B8. Equation (13) was used for all of the calculations in row 2 [3].

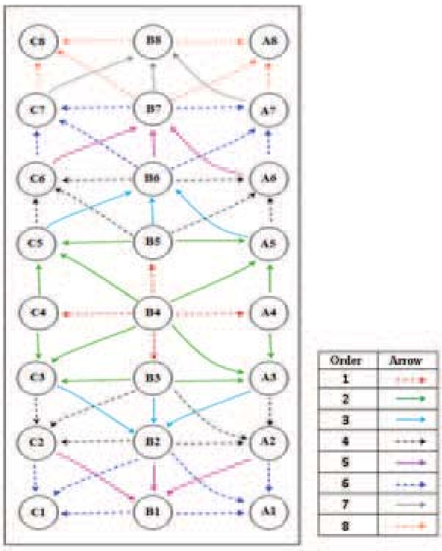

Fig. 2. Computation order of all sensors beginning at B4

III. EXPERIMENTAL SETUPS

An array of physical sensors was utilized to collect real solar irradiance data. This section shows the implementation of the sensors and the array. The experiment employed was developed from those used in previous studies [5][6]. The sensor array used in this study was designed to have the same physical size as 24 Kyocera KD135GX 125W panels set-up in a traditional series-parallel configuration. These panels were placed on adjustable legs so that they could collect data at 0° and 26°, the latter being the pitch of a typical household roof [2]. The sensor array had the dimensions of a 17.5 ft x 10 ft steel structure (see Fig.3). Sensor boxes placed on the steel base were distributed in a matrix form that began at with A1 and ended at C8 (see Fig.3). This arrangement was used so that each sensor could be accessed easily. The separating distance between the sensors was 2.5 ft (horizontally) and 5 ft (vertically). All of the solar cells were identical; each was glued directly to the box lid. The solar cells used in this study were supplied by the Solar Made company [7]. Each sensor was rated at 0.5V 125mA. The area of each cell was approximately $5cm^2$ (2.5cm x 2cm), as illustrated in Fig. 4. All of the sensors, as well as the solar cells, were placed in plastic boxes with removable lids to protect them from damage and dust [3].

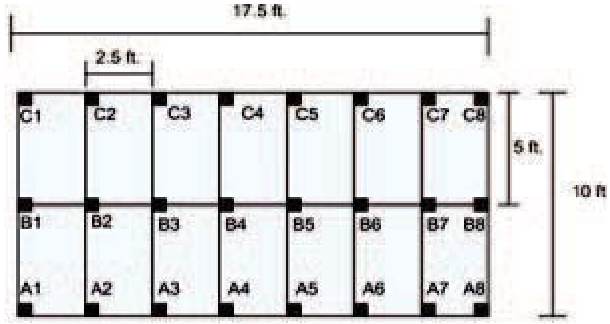

Fig. 3. Sensors (black boxes) Distributed around the panel

Fig. 4. Solar cell sensor [3]

A CAT5e cable was used to transfer the read data (by the sensor) to the main data acquisition systems (DAQs). This data was stored in a 1GB compact memory card inserted into each DAQ. Figure 5 is a photo of the single DAQ, (with eight inputs) that was used in this study. The DAQs were R-Engine-A programmable controllers manufactured by Tern, Inc. [8]. Three DAQs were used to cover all twenty-four data sources because each DAQ had eight imports [3].

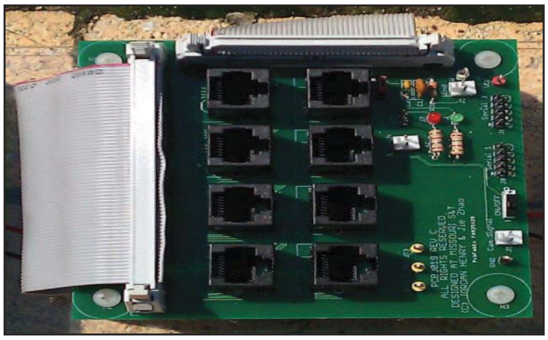

Fig. 5. Single data acquisition system with eight inputs [3]

The DAQs were placed inside an enclosed suitcase where they ran continuously to obtain data (load current, short circuit current, open circuit voltage, time, and date) from each sensor every other second. A very low watt light bulb was used to heat the DAQs slightly during the winter when they stopped working at temperatures below 40° F (as indicated in their specifications). The stored data in the compact memory cards was saved in Microsoft Excel table format. The memories were replaced by empty ones every week to ensure that the DAQs were still running. A straightforward analysis was conducted on and graphs were created from the data entered in Microsoft Excel. However, for more accuracy and higher degree of freedom in data analysis, MATLAB software was used [3] to perform detailed analysis and to write the final code to achieve the main objective of this paper.

IV. DATA CALIBRATION AND AVERAGING

A USB memory adapter was used to collect and import all readings from the 24 sensors from the DAQ system to MATLAB software. The array was located in a place that was open to sunlight; it was laid flat most of the day. Two main statistical processes were applied before the data was worked on. Data calibration was completed first to ensure that all of the solar cells were identical. This process was conducted by multiplying all data by such a factor to meet them up at a reference sensor reading as in (14).

$$i_{short(new)} = i_{short(old)} \times \frac{\left(i_{ref.(m)}/i_{old(m)}\right)+\left(i_{ref.(n)}/i_{old(n)}\right)}{2} \quad (14)$$

where $i_{short(new)}$ is the new short circuit current to be calculated, $i_{short(old)}$ is the old short circuit current to be calibrated, i_{ref} is the reference current, and m and n were two points through the day. These points were chosen so that the calibration factor was, approximately, the average among all factors throughout the day. (One calibration factor was tested. This calibration provided results that were less efficient than those provided by two-point calibration.) The data was averaged over one minute. Figure 7 illustrates the initial process (data calibration) that preceded the main analytical processes.

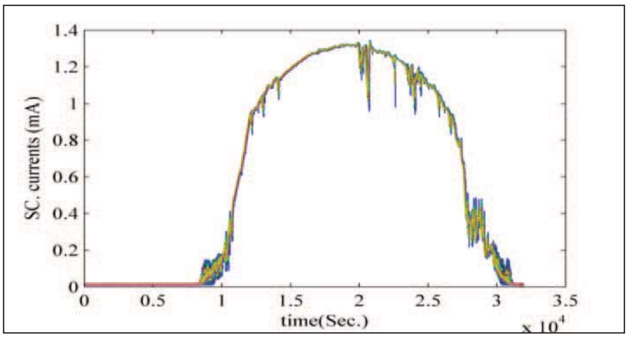

Fig. 6. Calibrated Sensors Readings

V. SOLAR RADIATION MODEL AND ACTUAL DATA FITTING

Figure 7 is the power versus time curve from the Meinel and Meinel model taken in April 1st. The time axis scale was converted to a minute scale rather than an hour scale so that the result could be compared with actual data.

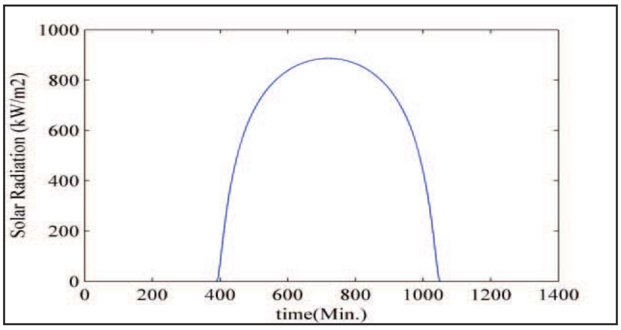

Fig. 7. Curve of the sun radiation on April 1 (created with the physical model)

The main equation used in the Meinel and Meinel model was adjusted to convert the y-axis of the sun radiation spectrum from (kW) to (mA) by such factor. This factor was the maximum reading that was taken from the actual data divided by the maximum value that was taken from the model. The relationship between the output power and the short circuit current of a solar cell was nearly linear [1]. An azimuth angle (Ψ=7.0°) was included in the model calculations. This process was completed so that the mathematical model could be matched with the actual data. The mathematic model was the plotted over the actual data so that it could be analyzed (See Figs. 9 and 10).

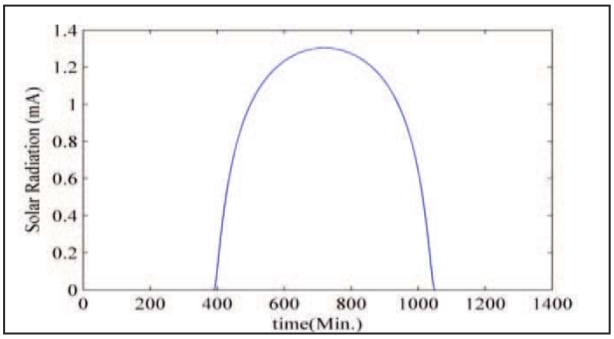

Fig. 8. Solar radiation on April 1

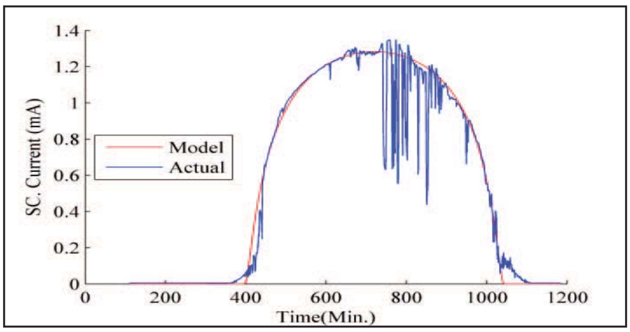

Fig. 9. Actual data and mathematical model for April 1

VI. ESTIMATING FUTURE TRENDS

The linear regression method was used to predict the difference between the actual data and the mathematical model for sun radiation. Both subtraction and division were attempted. Figure 10 illustrates a comparison between the two. Subtracting the mathematical model from the actual data (given in Fig. 11) provided a smaller mean squared error than division provided. Equation (7) was used to apply linear regression on the subtracted data. The predicted signal was then added to the Meinel and Meinel model for the same day so that the predicted sun radiation could be obtained (see Figs11 and 12).

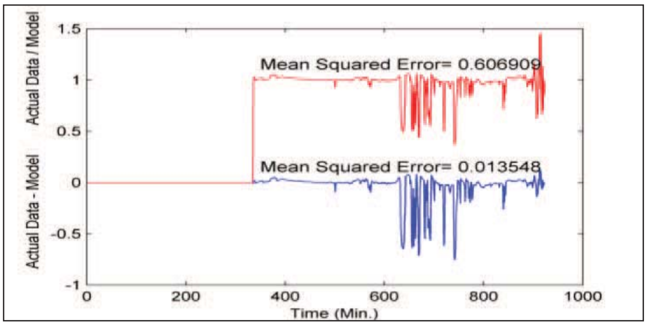

Fig. 10. Comparison between subtracting and dividing

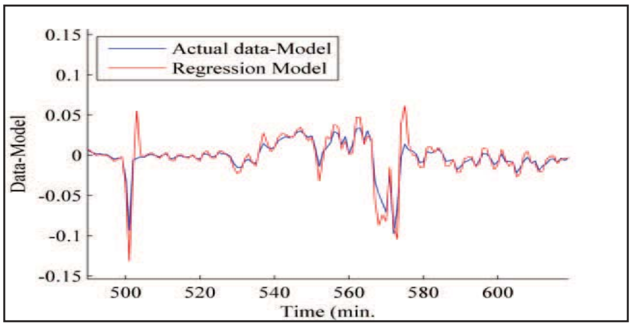

Fig. 11. Applying the linear regression method

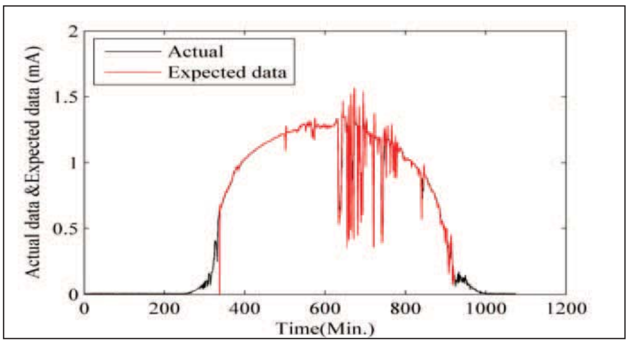

Fig. 12. Adding the predicted signal to both the Meinel and Meinel model as well as the actual data for April 1

Additional comparisons were made to ensure that subtraction was better than division. These additional comparisons were also made to select the best window size for the linear regression equations. Four different values of window size (3, 4, 5, and 6) were tested. A comparison between subtracting and dividing the mathematical model with the predicted data, after linear regression was completed for different window sizes for four days is summarized in Table I. The results indicate that applying linear regression to subtract the mathematical model from the actual data, with a window size, equals four yields better than the mean squared error.

TABLE I. A COMPARISON BETWEEN DIFFERENT WINDOW SIZES

DAY	*MSE % (SUBTRACTING) (WINDOW SIZE)*				*MSE % (DIVIDING) (WINDOW SIZE)*			
	3	4	5	6	3	4	5	6
April 1	3.02	2.70	2.94	3.20	3.08	2.88	3.18	3.56
April 2	3.09	2.11	2.43	3.70	2.89	2.97	2.99	3.07
April 3	2.92	3.01	3.37	3.90	3.66	3.02	3.20	3.09
April 4	3.91	2.39	2.89	4.11	3.76	4.06	4.00	3.80
Mean	3.23	2.50	2.90	3.70	3.30	3.23	3.30	3.30

A. Data Filtering and Noise Rejection

Solar radiation during a partly cloudy weather condition goes higher than the normal peak at sunny day as explained in [9].Therefore, the predicting signal may exceed the sun light limitation due to the sharp change in the solar radiation. Accordingly, a high radiation filter (HRF) was used to reject the unacceptable values. Readings that were higher than 20% of the values calculated from the Meinel and Meinel model were rejected when the HRF was being designed. These sensors also provided strange readings. For example, great deal of noise at sunrise and sunset. To fix that, approximately 100 readings were clipped minutes after sunrise and before sunset.

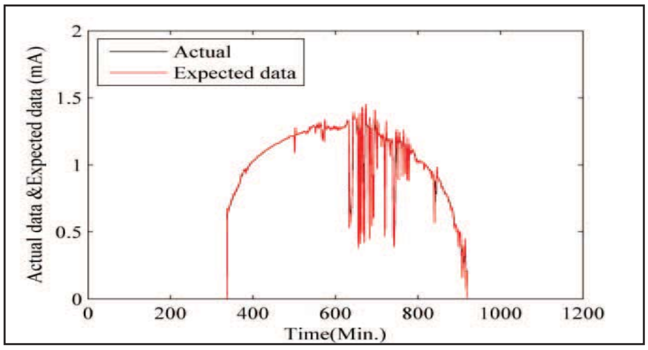

Fig. 13. Data after filtering and noise Rejection

B. Array Data Construction

Generating all of the sensors' readings from only one sensor reduced the costs of both control circuits and DAQ systems. It also sped up the data analyzing with a smaller room of data storage. All of the sensor readings could have been created from sensor B4 (the center of the array) (see Fig. 2). This work extended to combining previous processes (data prediction and generation) into one step. Figures 14 and 15 compare all array actual data and sensors data after predicting B4 sensor. They also generate the other 23 sensors' data using linear regression method, slopes among the sensors and correlation factors.

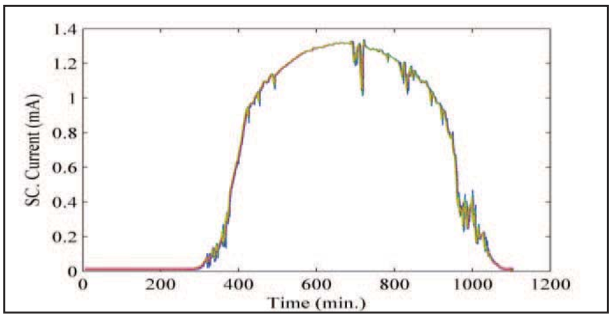

Fig. 14. Actual data for April 3

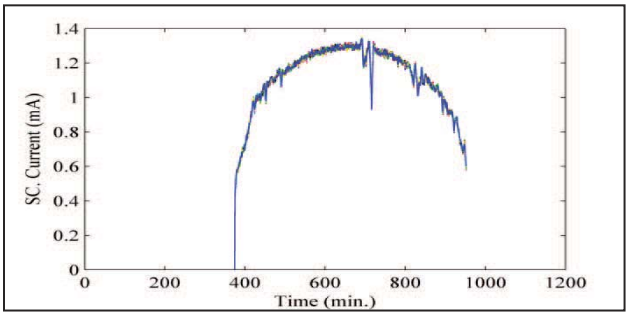

Fig. 15. Predicted data for April 3

The MSE was identified for each individual sensor. The average error was calculated at 2.6%. The difference between the predicted data, the actual data, and the MSE for sensor A1 is illustrated in Fig. 16.

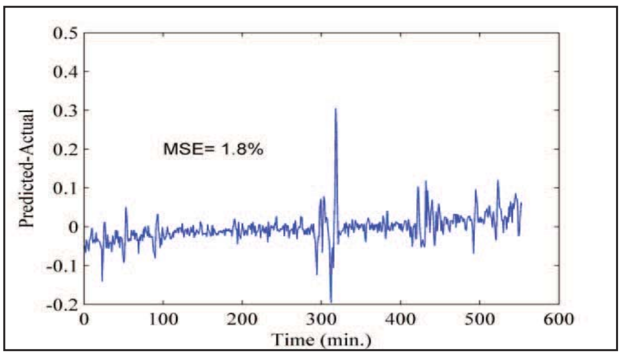

Fig. 16. MSE error for sensor A1

VII. CONCLUSIONS

The Meinel and Meinel model, along with the linear regression method, were used to predict the future trend of the solar radiation yields better results. The best results obtained using a window size equals three for the regression equations. The results from this study indicate that different parts of a solar panel can be related to one another. Additionally, to predict the next minute reading of each part of a solar panel using both model fitting and the correlation method provided an acceptable result, with an average MSE of 2.6%. This technique reduces not only the time needed to conduct data analysis but also the space needed for data storage. This technique also reduces the costs of control circuits and DAQ systems. Sun radiation during cloudy weather may exceed the maximum limit (reaching 1.07 p.u) because the clouds reflect this radiation. This condition may have a negative effect on data prediction. Therefore, a high radiation filter (HRF) should be used to correct the calculated radiation at the peak time.

ACKNOWLEDGMENT

This work was supported, in part, by the National Science Foundation under award ECCS-0900940. We would like to thank King Saud University, Riyadh, Saudi Arabia, Najran University, Najran, Saudi Arabia, and Saudi Arabian Cultural Mission, Washington D.C., for their support.

REFERENCES

[1] J. Zeng; W. Qiao, "Short-term solar power prediction using an RBF neural network," Power and Energy General Meeting, 2011 IEEE

[2] R. Messanger, J. Venture, "Photovoltaic Systems Engineering," 2nd ed., New York. CRC Press, 2004

[3] F, Alfaris, "Stochastic model for solar sensor array data," M.S.Thesis, ECE., Missouri S&T., Rolla

[4] X. Yan and X. Su, Linear regression analysis. theory and computing. New Yok. World Scientific, 2009, pp. 201–204.

[5] B. A. Yount, "MANAGING A SOLAR SENSOR ARRAY PROJECT. ANALYZING INSOLATION & MOTIVATION," Master of Science, Engineering Managment, Missouri University of Science and Technology, 2011.

[6] N. Nagarajan, "Development of a graphical user interface for the study of parallel-connected solar array," Missouri University of Science and Technology, 2011.

[7] SolarMade Company [online]. Available. http.//www.solarmade.com.

[8] Tern Inc [online]. Available. http.//www.tern.com.

[9] Nijhuis, M.; Rawn, B.G.; Gibescu, M., "Classification technique to quantify the significance of partly cloudy conditions for reserve requirements due to photovoltaic plants," PowerTech, 2011 IEEE Trondheim , vol., no., pp.1,7, 19-23 June 2011