3rd International Conference on Engineering and ICT (ICEI2012)
Melaka, Malaysia
4 – 6 April 2012

SOLAR INSOLATION FORECAST USING ARTIFICIAL NEURAL NETWORK FOR MALAYSIAN WEATHER

H.G. CHUA[1,a], B.C. KOK[2,b] and H.H. GOH[3,c]
[1,2,3] Universiti Tun Hussein Onn Malaysia
Department of Electrical Power Engineering
Faculty of Electrical and Electronic Engineering
86400, ParitRaja, BatuPahat, Malaysia
[a]Email: johnathanchua_86@hotmail.com , [b]Email bckok@uthm.edu.my, [c]Email: hhgoh@uthm.edu.my.

***Abstract*—Solar insolation forecast is essential for photovoltaic (PV) generation plant in enhancing the usage of solar energy for electrical production scheme. Likewise, it improves the PV power generation efficiency by regulating the control algorithm and charge controller corresponding to the prediction probability. An acquired data in solar insolation is required for solar energy harvesting. This paper presents the 12 hourly solar insolation forecast using Artificial Neural Network (ANN). A Multi-level perceptron (MLP) with back propagation technique model is proposed to predict the next day 12 hours solar insolation. The performance of MLP model is investigated with 60 days of solar insolation data from July 9th to September 9th 2011. The investigation is conducted under two different tropical weather conditions, sunny and rainy conditions. Hence, the best performance MLP forecaster model with a minimal error is selected through a trial and error method under various weather conditions. In this paper, the performance of the forecaster is shown. The results allow inferring the adequate performance and pertinence of this methodology to predict complex phenomena, such as solar insolation.**

***Keywords*— Solar Insolation, ANN, MLP, Forecasting, Solar Energy Harvesting.**

I. INTRODUCTION

There has been a continuous demand of electrical energy to support a world community needs as whole. Without electrical energy supplies, the entire economic activities are dysfunctional as the demand of necessities are not met. In most countries, natural resources such as petroleum and fossil fuel are the prime sources for their electrical energy production. The dependency on natural resources is still high as the sources gradually decreases per year leads to a spike in market price. Studies and research has been taken up to curb the awareness of this issue by replacing fossil fuel with other renewable energy and one of it is solar energy.

Malaysia is one of the tropical country which lies entirely on the equatorial region. The solar insolation receives in Malaysia is averagely ranging from 4.21 kWh/m^2 to 5.56 kWh/m^2 [1]. Whereas the ambient temperature remains uniformly throughout the year as its average scale ranges between 26 °C to 32 °C [1]. Thus, the abundance sunshine throughout the year receives in Malaysia spurs the interest of harnessing energy.

A detailed of solar insolation data is required as it designates an interest for potential location with the highest solar energy measurement [2] which enhances the research in energy utilisation [3] and application [4]. Unfortunately, due to the lacking of meteorological centres in rural areas, these solar insolation data is difficult to obtain because of cost and difficulty in measurement. Therefore, an alternative method on generating the forecasted data is constructed [3-5]. General practical approaches introduces in solar insolation modelling is forecasting and prediction. With the aid of this method, the operation control and the energy optimisation on harnessing the suns energy are boosted [4]. Artificial intelligence (AI) is a familiar successful method applied in modelling and forecasting the solar insolation data. The advantages of using the artificial neural network (ANN) application are it requires less knowledge of internal system parameters, involves less computational effort and the capability in solving multivariable problems [5]. Hence, ANN is effortless applied by most researches on this application in order to study the change of solar insolation throughout a potential location on solar energy.

ANN has various types of network topology and Multilayer Perceptron (MLP) is one of them. MLP has been used in many solar insolation forecast [5], likewise it is the cheapest method that uses a certain parameters such as latitude, longitude, altitude and sunshine duration [2], sunshine ratio [3], mean daily solar irradiance, mean daily air temperature, day and month [4], sunshine accuracy and mean average temperature [5], atmospheric data and insolation data [6]. Multilayer Perceptron (MLP) with Back propagation training algorithm topology is commonly used in most insolation level forecast [2], [3], [4], [5], [7], [8], [9], [10]. Some uses Radial Basis Function (RBF) topology which offers a better result than MLP restrictedly if there are more input parameters [2-3].

In this paper, MLP with back propagation is used to forecast 12 hour solar insolation based on two parameters i.e. solar insolation and time. The MLP forecaster is tested in two tropical weather conditions sunny and rainy. The best modelling is then selected using trial and error method. All simulation is done in MATLAB environment to assist the selection of best network modelling forecast.

II. LITERATURE REVIEW AND PREVIOUS WORKS

There are various ways on ANN implementation into solar insolation forecast. This section explains the ANN implementation used by previous authors based on the network topology, training algorithm, network generalisation capability and the forecasting application.

M.Mohandes, A.Balghonaim, M.Kassas, S.Rehman and O.Halawani (2000), uses radial basis function (RBF) network to develop monthly mean daily solar radiation falling on the horizontal surface and compare the performance network with multilayer perceptron network and a classical regression model. The proposed input network uses latitude, longitude, altitude and sunshine duration from 41 different locations. An extra input is included, indicating the number of months into RBF network. The results shows that RBF outperformed MLP, as the average mean absolute percentage error (*MAPE*) for RBF is 10.1, lower than MLP, 12.6. Moreover, empirical regression model such as Angstrom (1924) and Rietveld (1978) does not surpass RBF as it gives a better result [2].

Atsu S.S. Dorlo, Joseph A. Jervase, Ali Al-Lawati, (2002), compare both forecaster using MLP and RBF network. The solar radiation is estimate using clearness index. The proposed input network is month of the year, latitude, longitude, altitude and sunshine ratio from eight different stations in Oman. Their results shows both network performed well based on root mean square error (*RMSE*) as RBF error ranges from 0.83 to 10.08 MJ/m2/day, while MLP error ranges from 1.01 to 9.41 MJ/m2/day. Consequently, they commented that the best MLP network uses 3 hidden layers due to minimal mean and standard deviation of the root mean square errors. The authors conclude that RBF is the best network due to less computation time [3].

Adel Mellit and Alessandro Massi Pavan, (2010), forecast 24 hours solar irradiance using MLP network based on mean daily solar irradiance, the mean daily air temperature and the day of the month as input parameters. The authors introduce the network generalisation capability improvement by using K-fold cross validation method. Four experiments are taken based on sunny and cloudy weather. The best MLP architecture obtained by them is with two hidden layers as the first contains 11 neurons and the second contains 17, $\{3\times11\times17\times24\}$. As their results, the coefficient of determination, (R^2), for sunny weather ranges from 0.95 to 0.99 while cloudy weather ranges from 0.92 to 0.97. Moreover, the authors compare the effectiveness of the forecast network with the grid connected photovoltaic plants which give a good correlation of determination result of 0.90. In addition, the result shows that the mean absolute error (*MAE*) is less than 5% and the correlation coefficient ranges from 90% to 92% [4].

Melit, Shaari, H.Mekki and N.Khorissi (2008), introduced a reconfigurable field programming gate array (FPGA) device such as (Xilinx and VirtexII) using VHDL to forecast the daily solar radiation based on ANN architecture. The design device is implemented for real time forecasting application. The network uses a backpropagation MLP and sunshine duration and mean average temperature as input parameter. The forecast simulation performance of VHDL is then compared with the simulation performance written in MATLAB program to determine the accuracy of the result. They simulated in MATLAB as they obtained the best architecture of a single layer with 9 neurons resulted 98% coefficient of determination, R^2. On the other hand, 18 bits fixed points use in VHDL gives an acceptable result for the forecast simulation. As their overall conclusion, they remark a good agreement between the data simulated from MATLAB and VHDL [5].

K.S Reddy and Manish Ranjan (2003), uses backpropagation multilayer feed forward network based model for estimation of monthly mean daily and hourly values of global radiation. They obtained better prediction results compare with other empirical regression model. In the network architecture design consist of an input layer, two hidden layer and an output layer. The first hidden layer uses seven neurons while the second hidden layer uses eight neurons. They used 9 parameters input to estimate the radiation. In their case study, the data set is divided into two regions in India, north and south, and based on three different seasons summer, rainy and winter. The mean absolute relative deviation (*MARD*) for their prediction is about 4%. The authors compare the prediction accuracy between two regions in India, New Delhi and Mangalore, based on summer, winter and rainy season. Therefore, the accuracy results of three predictions of the maximum absolute error (*MAE*) are minimal [7].

Mónica Bocco, Gustavo Ovando and Silvina Sayago (2006) developed a backpropagation type neural network to estimate the solar radiation based on five types of input parameters. The author uses three layer network and five neurons in the hidden layer. They simulated 8 types of forecast results from models, M1-M8. Model of M6-M8 are records of cloudiness at 14 hours shows an average error in estimation. Moreover, model M3 and M4 are based on temperature and precipitation information which shows the greater errors in the estimation. The best estimation result models are M1 and M2. M1 is analysed through a dispersion diagram of observed and estimated value. Through the result they observed, the underestimation percentage reaches 15% when the daily radiation exceeds 25MJ m^{-2} d^{-1}. In their overall models, the estimation indicates the *RMSE* between 3.15 and 3.88 MJ m^{-2} d^{-1} [8].

Özgür, Humar, Ali and Muammer (2010), proposed a standard backpropagation and backpropagation with momentum training algorithm in daily solar radiation prediction. The network topology consist of 5 input parameters such as latitude, longitude, altitude, day of the year and mean temperature, a 10 neurons hidden layer and an output. Therefore from their result, number of iteration and coefficient of determination, R^2 from both standard and momentum backpropagation are 15000, 0.9870 and 7500, 0.9821 as momentum backpropagation compute faster with better correlation. On the other hand, the *MAE* and *RMSE* result of a standard backpropagation outpaced the momentum backpropagation by 1.02% and 0.8% difference. As their conclusion overall, the standard backpropagation gives better result with slow computational time than the momentum backpropagation as the mean relative error concluded with 8.96% and 10.12% each [9].

Adnan Sözen, Erol Arcaklioğlu and Mehmet Özalp (2004), studied two cases on solar potential. The first study is to estimate the solar potential using ANN based on metrological and geographical data. Then the second study is to find the best estimation results for each station using different learning algorithm and a logistic sigmoid transfer function. Three years of metrological data from 17 different stations is taken to train the neural network. The author uses 6 parameters as an input layer of the network. Therefore as the result, the maximum and minimum mean absolute percentage error in testing data from two cities of Muğla and Antalya are 6.735% and 2.921%. Meanwhile, the absolute fraction of variance, R^2 of two cities are 99.3% and 99.893% respectively. The minimum mean absolute percentage error and R^2 values for training data is 1.623% and 99.9658% [10].

III. DATA COLLECTION

In-house measurement data such as solar insolation of 2 mounted photovoltaic (PV) panels position oppositely to each other are taken within 2 months, 60 days from July 9th 2011 to September 9th 2011. Fig. 1 shows the PV system architecture where the solar insolation measurements are taken at each solar panel. Two solar insolation data logger equipment (*COM 3 and COM 4*) are used and it is tilted the same angle as the PV generator. Due to Malaysia lies entirely in the equatorial region, only 12 hours data from 7am to 7pm under various climate conditions is taken as this period the sun radiates the most. Fig. 2. and Fig. 3. show the experimental average 12 hours daily solar insolation per day and per hour.

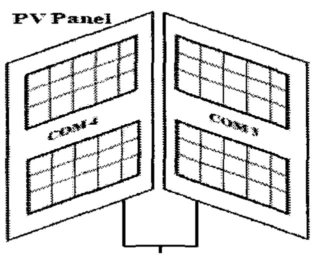

Figure 1. Solar insolation data logging using (***COM 3 and COM 4***)

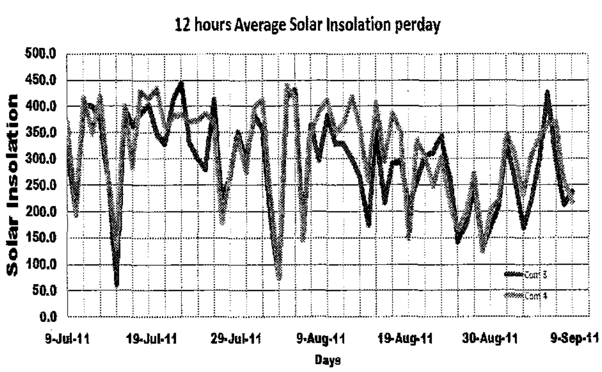

Figure 2. Average of Solar Insolation per day

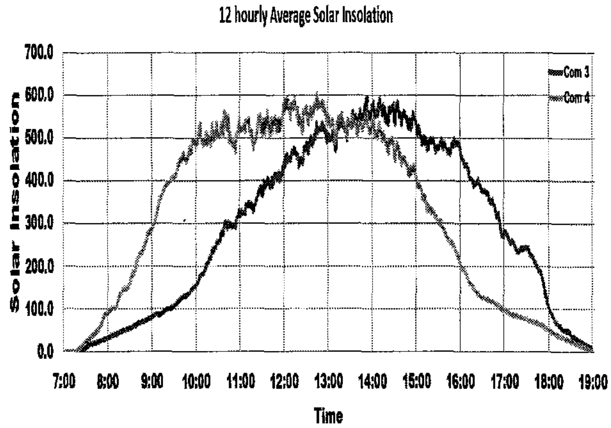

Figure 3. Average of Solar insolation per hour

IV. ARTIFICIAL NEURAL NETWORK (ANN)

ANN is applied in many areas and one of the common application areas applied is modelling and forecasting. This application curbs the many complexes problem for conventional computers and human beings [10]. Each function of the ANN imitates the four important biological operation of the human brain. These operations receive input, combine them and produce an output result [4].

ANN is classified by it network structure, training or learning algorithm and activation function. The structure comprise of input layer, a few hidden layer and an output layer [10]. One of the most common structures used is feed-forward network [4]. Each network has different styles of connection and learning algorithms which are fitted suitably to its architecture and convergence time.

There are two types of learning algorithm, supervised and unsupervised. With supervised learning, the weights are adjusted for the network to produce the desired output. On the other hand, unsupervised learning does not depend on the target data to achieve the desired output. The aim is to find the core structure of the data. One of the common supervised learning algorithms used is back-propagation algorithm, which has different variants [10].

The activation function refers to the output relation of the network to the input based on the input activity level. Sigmoid function is widely used as activation function due to its non-linearity function whose output lies in between 0 and 1 [7]. Thus from Mellit, S.Shaari, H.Mekki and N.Khorissi comments, a majority uses multi-layer perceptron (MLP) with a back-propagation (BP) training algorithm in most application [5]. Likewise, in this paper it aims to use the similar method.

V. SOLAR INSOLATION FORECASTING USING MLP NETWORK

In most research, multi-layer perceptron (MLP) network is used in modelling and forecasting [5] due to its well-known feed-forward structure [4]. In this paper, the MLP structure comprises of an input, output and a hidden layers. This structure imitates the basic function of the human brain as it receives inputs, combine them and produce final output result [4]. The input data are divided into training, validation and test sets. The input and output data are normalised in the range between -1 and 1. Fig. 4. shows the MLP input parameter such as time and solar insolation, while the output produce the next day 12 hours solar insolation.

MLP network has various connection styles and learning algorithms as it is adapted to its structure and convergence time. Back-propagation is a popular supervised learning algorithm [10] and it is used in this research due to its ability to adjust the weights for the network in producing the desired output. Without supervised learning algorithm, the desired output is unachievable as the weights are not adjusted to the actual target data.

Activation function is used in between each hidden layers. Hyperbolic tangent sigmoid (*Tansig*) activation function is used at input-hidden layer whereas linear transfer function (*Purelin*) activation function is used at hidden-output layer.

MLP network is simulated in MATLAB environment. The Levenberg-Marquardt algorithm is used as a numerical tool to minimise the error during training. The best MLP structure lies on choosing the best activation function and number of neurons in the hidden layer. Trial and error method determines the results of a suitable number of neuron in each model. The designs steps of the MLP architecture are explain in flowchart as shown in Fig. 5.

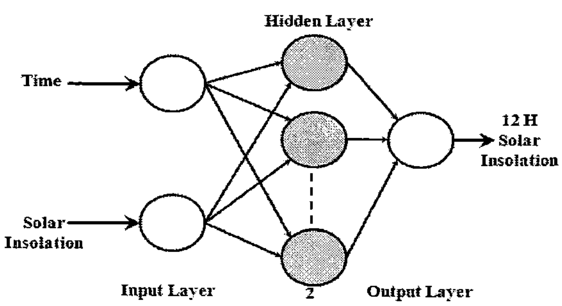

Figure 4. MLP network for 12-Hours forecast

Figure 5. Block diagram of MLP System design and architecture

VI. SIMULATION RESULTS OF MLP FORECASTER

The best structure for the MLP forecaster is selected through the best activation function and number of neurons in the hidden layer use in the network. Activation functions are used in between each hidden layers. While constructing the network, users would have to choose the suitable activation function which is defined in MATLAB. "***Tansig***" activation function is used at input-hidden layer whereas "***Purelin***" activation function is used at hidden-output layer. According to analysis done by M. Badrul, "***Tansig***" and "***Purelin***" are the best structure due to minimal mean square error (MSE) [11]. Fig. 6. depicts the activation function used in the MLP forecaster network.

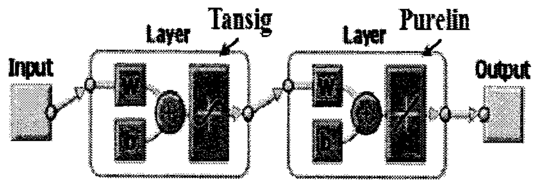

Figure 6. Activation function used in MLP network

The forecasting results in two different weather conditions i.e. during sunny and rainy on July 11th (2011) and July 15th (2011) are shown in Table I. The correlation coefficient, r indicates how close the predicted and measured data is. Fig. 7. shows the scatter plots of correlation coefficient for sunny day. *MSE* (Mean Square error) provides information on long-term model performance which specifies the average deviation between the predicted values to the corresponding measure values. When the coefficient of determination R^2 approaches 1 and *MSE* approaches zero this signifies the solution of the problem provides the most accurate answer [9]. Table I shows the prediction results with highlighted bold minimal error.

Table I. Forecasting results of two loggers and two weather conditions July 11th (Sunny) and July 15th (Rainy)

No. of Nodes	Com 3						Com 4					
	Sunny			Rainy			Sunny			Rainy		
	R2	MSE	r	R2	MSE	r	R2	MSE	r	R2	MSE	r
1	0.994	0.003	0.997	0.921	0.0075	0.974	0.979	0.012	0.99	0.966	0.0094	0.973
2	0.978	0.01	0.991	0.905	0.0092	0.953	0.989	0.005	0.996	0.952	0.0088	0.974
3	0.986	0.006	0.993	0.956	0.0044	0.981	0.991	0.005	0.997	0.967	0.0067	0.981
4	0.988	0.006	0.994	0.942	0.0064	0.977	0.992	0.004	0.996	0.973	0.0045	0.992
5	0.982	0.013	0.984	0.921	0.0076	0.967	0.981	0.009	0.996	0.969	0.0059	0.985
6	0.977	0.01	0.997	0.968	0.0031	0.988	0.995	**0.002**	0.998	0.946	0.0092	0.982
7	0.978	0.009	0.998	0.963	0.0038	0.991	0.982	0.008	0.992	0.943	0.0098	0.978
8	0.976	0.01	0.993	0.948	0.0049	0.984	0.993	0.003	0.997	0.998	**0.0005**	0.999
9	0.997	**0.002**	0.999	0.959	0.0042	0.989	0.983	0.008	0.993	0.997	0.001	0.997
10	0.984	0.006	0.996	0.92	0.008	0.963	0.995	0.003	0.999	0.979	0.0036	0.99
15	0.977	0.01	0.996	0.993	**0.0009**	0.997	0.992	0.004	0.997	0.968	0.0055	0.988
20	0.978	0.01	0.995	0.98	0.0023	0.992	0.985	0.008	0.997	0.976	0.0041	0.992
25	0.982	0.008	0.993	0.945	0.006	0.987	0.996	0.003	0.998	1	0.0015	0.996
30	0.985	0.007	0.997	0.942	0.0067	0.983	0.992	0.006	0.996	1	0.0024	0.993

A summary of determined R^2, *MSE* and *r* values are given in Table I. According to the sunny weather prediction results in section *COM 3*, the R^2 varies in the range of 97.60% - 99.70%, *r* is between 98.40% - 99.90% and MSE varies from 0.2% - 1%. Whereas for section *COM 4*, R^2 varies in the range of 97.90% - 99.60%, r is between 99.00% - 99.90% and *MSE* varies from 0.2% - 1.2%. From the rainy prediction results observation in section *COM 3*, the R^2 varies in the range of 90.50% - 99.30%, *r* is between 95.30% - 99.7% and *MSE* varies from 0.09% - 0.92%. As for *COM 4*, R^2 varies in the range of 94.30% -1%, r is between 97.30% - 99.9% and *MSE* varies from 0.05% - 0.98%

Fig. 8. and Fig. 9. illustrate the two loggers prediction results for sunny day on July 11th 2011 while Fig. 10. and Fig. 11. depict the two loggers prediction results for rainy day on July 15th 2011.

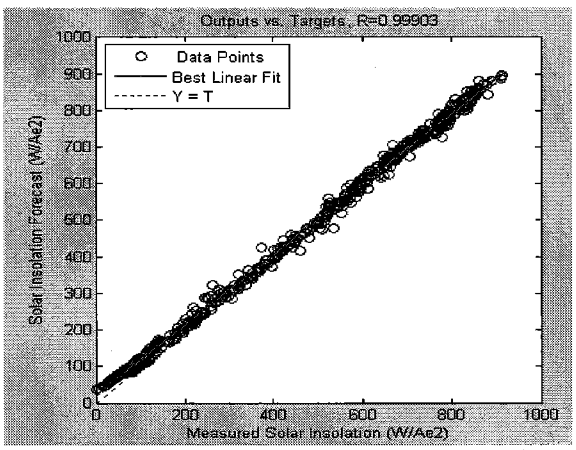

Figure 7. Coefficient correlation of solar insolation for sunny day on July 11th 2011 using com 3 logger.

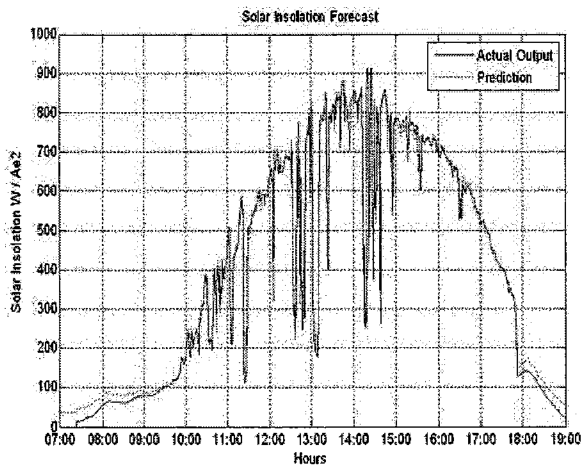

Figure 8. (Com 3) Solar insolation measured and prediction on July 11th 2011 (Sunny day) using MLP model 9.

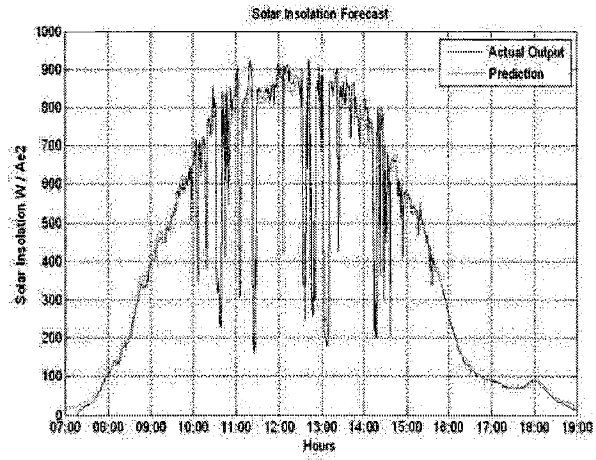

Figure 9. (*Com 4*) Solar insolation measured and prediction on July 11th 2011 (Sunny day) using MLP model 6.

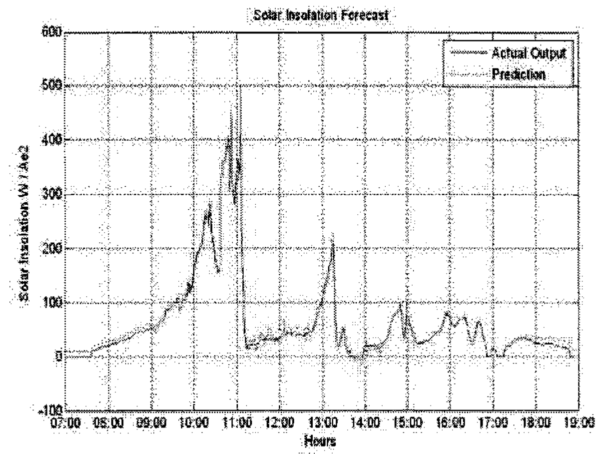

Figure 10. (*Com 3*) Solar insolation measured and prediction on July 15th 2011 (Rainy day) using MLP model 15.

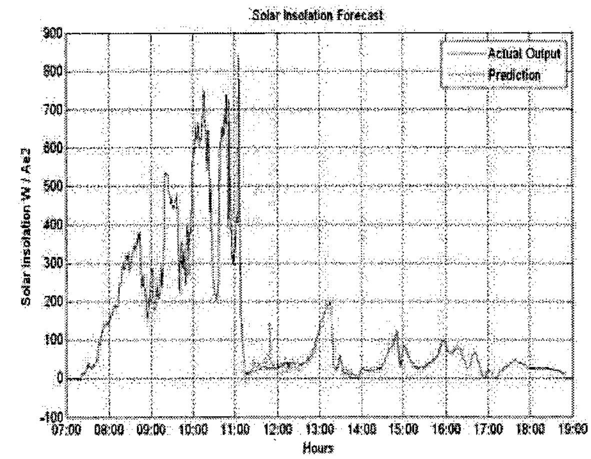

Figure 11. (*Com 4*) Solar insolation measured and prediction on July 15th 2011 (Rainy day) using MLP model 8.

VII. DISCUSSION

The results and analysis of the forecaster undergo two types of weather conditions. Minimal error for sunny and rainy weather conditions for Com 3 shown in Table I are network with 9 nodes and 15 nodes, respectively. The values of MSE computed from the network are 0.2% and 0.09% each. Whereas the minimal MSE value for Com 4, is the network with 6 nodes for sunny weather and 8 nodes for rainy weather with the value error of 0.2% and 0.05% each. In both studies, the MSE is small due to fewer input parameter used into the network. The r and R^2, of the network is approximately reaching to 1. This indicates that the network do not overestimates as the predicted and measured data are closely adjacent. Although less parameter is used, this does not affect the performance of the forecaster to predict the next day 12 hours solar insolation.

VIII. CONCLUSION

This paper has presented a MLP network for 12 hours solar insolation forecast. ANN has been chosen due to its versatile and commonly used to forecast the solar insolation for any areas provided a comprehensive meteorological data is available. This forecasting process helps the effectiveness to estimate solar insolation, when only a limited number of meteorological variables are available. Although it has less input parameters, it has no adverse effect on the performance of the forecaster and this method can be used in future application with more input parameters. As a conclusion, this method can be a useful in any possible application regarding to optimal power tracker for stand-alone PV system. The method can be included into the PV system control algorithm to track power optimally corresponding to the solar insolation pattern. Therefore, this method helps in effective future planning and operation of PV system for the utilisation of solar energy harvesting.

ACKNOWLEDGMENT

The authors gratefully acknowledged the financial support by the Ministry of Higher Education (MOHE), Malaysia under Fundamental Research Grant Scheme (FRGS) for this research. The authors would also like to thank the staffs in Basic Electrical Machines Laboratory and ORICC, UTHM, for their support and valuable contributions towards the success of this research.

REFERENCES

[1] AyuWaziraAzahari, KamaruzzamanSopian, AzamiZaharim and Mohd Al Ghoul, 2008. A New Approach for Predicting Solar Radiation in Tropical Environmentusing Satellite Images-Case Study of Malaysia. WSEAS Transactions on Environment and Development, Issue 4 Vol.4, April 2008, pp. 373-378.

[2] M.Mohandes, A.Balghonaim, M.Kassas, S.Rehman and O.Halawani, 2000. Use Radial Basis Function for Estimating Monthly Mean Daily Solar Radiation. Elsevier Science: Solar Energy Vol.68, No.2, pp. 161-168.

[3] Atsu S.S. Dorlo, Joseph A. Jervase, Ali Al-Lawati, 2002. Solar Radiation Estimation using Artificial Neural Networks. Elsevier Science: Applied Energy, pp. 307-319.

[4] Adel Mellit and Alessandro Massi Pavan, 2010. A 24-h Forecast of Solar Irrandiance using Artificial Neural Network: Application for Performance Prediction of a Grid-connected PV Plant at Trieste, Italy. Elsevier Science: Solar Energy, pp. 807-821.

[5] Mellit, S.Shaari, H.Mekki and N.Khorissi, 2008. FPGA-based Aritificial Neural Network for Prediction of Solar Radiation Data from Sunshine Duration and Air Temperature. Region 8 Sibircon, IEEE, 2008, pp. 118-123.

[6] Y. Kemmoku, S. Orita, S. Nakagawa and T. Sakakibara, 1999. Daily Insolation Forecasting Using A Multi-Stage Neural Network. Elsevier Science: Solar Energy Vol. 66, No. 3, pp.193-199.

[7] K. S. Reddy and Manish Ranjan, 2003. Solar resources estimation using artificial neural networks and comparison with other correlation models. Elsevier Science: Energy Conversion and Management 44, pp. 2519-2530.

[8] Mónica Bocco, Gustavo Ovando and Silvina Sayago, 2006. Development and evaluation of neural network models to estimate daily solar radiation at córdoba, Argentina. Pesq.agropec.bras, Brasilia, V.41, n.2, pp.179-184.

[9] Özgür, Humar, Ali and Muammer, (2010). Prediction of Daily Solar Radiation using ANNS for Selected Provinces in Turkey. International Scientific Conference 19-20 November, 2010, Gabrovo, pp. 450-456.

[10] Adnan Sözen, Erol Arcaklioğlu, Mehmet Özalp, 2004. Estimation of solar potential in Turkey by artificial neural network using meterological and geographical data. Elsevier Science: Energy Conversion and Management 44, pp. 3033-3052.

[11] Mohd Badrul Hadi Che Omar, 2008. Estimation of economic dispatch (line losses) at generation side using artificial neural network. Thesis, Universiti Tun Hussein Onn Malaysia.