Recovery of the Solar Irradiance Data using Artificial Neural Network

Kah-Ching Ho¹, Boon-Han Lim*,¹,²,³, An-Chow Lai¹
¹Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Bandar Sungai Long, 43000 Kajang, Selangor, Malaysia
²Centre for Photonics and Advanced Materials Research, Universiti Tunku Abdul Rahman, Malaysia.
³Centre for Power and Electricity, Universiti Tunku Abdul Rahman, Malaysia.
E-mail: limbhan@utar.edu.my

Abstract. Global solar irradiance (GHI) data plays a major role in the design, performance assessment, and monitoring of a solar energy conversion system. However, data loss happens sometimes due to various reasons. Hence, it is important to recover the lost data. Unfortunately, due to the high cost of measurement devices, other meteorological parameters are often not available to aid the data recovery. Nevertheless, a photovoltaic (PV) system could be installed nearby the site which the output power is recorded. This paper explored the capability of using output power from a PV panel to predict the lost or unavailable GHI data in the past via the Artificial Neural Network (ANN) technique. 1 day, 10 days, and 30 days of GHI data, time, and output power of a solar panel in July 2020 were used to train three ANN models. The ANN models were then used to predict one day and thirty days GHI data in July 2019 based on the PV output power of that period. The result shows that the ANN model with a higher number of training data can predict the GHI data with lower statistical errors. The results prove that PV output power can be used to recover the unavailable GHI data in the past with acceptable accuracy. This technique is useful for the industry to recover the historical time series of GHI for designing a new renewable energy project or assessing the past performance of a renewable energy project.

1. Introduction
Solar irradiance, the quantification of the available power density of the solar energy, plays a major role in determining the design parameters and performance of the operation of energy conversion systems such as photovoltaic (PV) systems [1]. A ground-mounted weather station equipped with pyranometers or a solar irradiance meter is frequently used to log solar irradiance. If the pyranometer is orientated horizontally, the measured solar irradiance is called Global Horizontal Solar Irradiance (GHI). However, solar irradiance data are often lost due to malfunction data logger, sensor error, faulty devices, or power failure. Hence, it is important to recover the lost information using either a statistical method or an artificial intelligent method. Data should be recovered because it can be used to assess the past performance of a PV plant or it can be compiled as a complete set of historical time series of GHI data which can be used to determine the size, the optimal orientation of solar panels, and the optimal inverter sizing ratio of a new PV project [2].

Artificial Neural Network (ANN) technique to predict a lost solar irradiance dataset was reported to have higher accuracy as compared to mathematical methods [3]–[5] because ANN can be trained from a ‘healthy’ dataset acting as input [6]–[7]. The variables used to train the ANN were meteorological and geographical parameters with the correlation coefficient of predicted value exceeded 90% [8]–[9]. However, in many developing countries like Malaysia, a weather station with a complete set of
meteorological parameters is scarce but there usually has a PV system installed nearby [10], in which the electricity generation data can be a useful input for the training of ANN models. Besides, the PV systems could be installed earlier than a solar irradiance meter in many projects [10] so it is worth estimating the earlier solar irradiance for a plant assessment purpose. In this research work, the recoverability of solar irradiance data was investigated using the electricity generation information of a nearby PV system. The recovered data is then compared with the original measured data to assess the performance of the ANN model. This method can provide a useful technique of recovery of historical time series of GHI data across many sites in a country in which the GHI data bank is very important for the determination of the aforementioned design parameters for future PV projects.

2. Methodology

2.1. Data collection

GHI was measured by a pyranometer placed on the rooftop of Universiti Tunku Abdul Rahman (UTAR), located in Bandar Sungai Long, Malaysia. The output power of a solar panel of a nearby PV system is recorded at 5 minutes intervals from sunrise to sunset via a microinverter which injects the solar power into the local grid. Complete GHI data without any loss and the output power of a PV panel were extracted from July 2019 and July 2020. The GHI and PV output power datasets in July 2020 were used to train ANN models whereas the GHI data in July 2019 were acted as a reference for comparison when ANN models were used to predict GHI data of July 2019 using the output power of a PV panel recorded in July 2019. The objective is to evaluate whether the information from 2020 can predict unavailable data in the previous year. Not only that, the use of the data from the same month, i.e., July, is to prevent any unforeseen variation of meteorological information such as seasonal change of weather and the sun path, that might affect the accuracy of the ANN prediction model. In the ANN analysis training, the time and the output power of the solar panel act as inputs, while the GHI acts as an output.

2.2. ANN analysis

The ANN model is made up of three layers [11], i.e. the input layer, the hidden layer, and the output layer as shown in Figure 1. The variables such as time and output power of PV panels are positioned in the input layer. For this analysis, the number of neurons is set to 10 in the hidden layer. The collected GHI data is positioned as the output layer.

![Figure 1. Schematic drawing of the ANN model used for solar irradiance training.](image)

The ANN model was developed using MATLAB 2016a. The ANN analysis for predicting GHI uses Levenberg-Marquardt backpropagation (LM) because this algorithm gives the most accurate solution as compared to the other algorithms such as gradient descent (GD), resilient backpropagation (RP), and scaled conjugate gradient (SCG) [7].

2.3. ANN models
The entire ANN analysis datasets are divided into two categories which are training data and testing set. The training set and validation set are subdivision category of the training data [12]. The training set is used to adjust the weights on the neural networks according to its error. The validation set is used to minimize overfitting which to measure the network generalization and it will halt the training if the generalization stops improving. The testing set provides an independent prediction of the network performance and will not affect the training of the ANN. For this research work, the training set, validation set, and testing set consist of 80%, 5%, and 15% of the sample data respectively. The percentages are distributed based on the knowledge provided in Ref. [12] and Ref. [13].

There are three models of ANN which were trained to predict the GHI. The first model only uses one-day data in July 2020 for the training and it is named as ANN model 1-day while the second model and third model use 10 days data and 30 days data, respectively. Therefore, they are named as ANN model 10-days and ANN model 30-days, respectively.

2.4. Performance of the prediction

A total of 30 days of GHI data was used to evaluate the performance of ANN models. The performance indicators are the percentage error of each day, the percentage error of a total of 30 days, and the root mean square error (RMSE). The percentage error and RMSE are expressed in Eqns. (1) and (2)

$$\% \text{ of daily error} = \left| \frac{\sum X_i - \sum Y_i}{\sum X_i} \right| \times 100 \quad (1)$$

$$\text{RMSE} = \left[ \frac{1}{N} \sum_{i=1}^{N} (X_i - Y_i)^2 \right]^{1/2} \quad (2)$$

where $N$ is the number of data, $X_i$ is the measured GHI value and $Y_i$ is the GHI value predicted by the ANN model.

3. Results and Discussions

3.1. Comparison of predicted GHI data using ANN models with original measured data

The predicted GHI data for 1 day and a total of 30 days in July 2019, using ANN models trained from data in July 2020, were compared with the original measured GHI data of July 2019 to assess the accuracy of the ANN model 1-day, ANN model 10-days, and ANN model 30-days. The results on 2nd July and 30 days are tabulated in Table 1 where ANN model 30-days has the best performance, i.e., the lowest RMSE value, followed by ANN model 10-days while ANN model 1-day shows the highest value for RMSE.

| Day      | RMSE (W/m²)       |
|----------|-------------------|
|          | ANN model 1-day   | ANN model 10-days | ANN model 30-days |
| 2nd July | 119.064           | 45.206            | 29.671             |
| 30 days  | 135.722           | 89.718            | 73.676             |

The percentages of daily error (see Figure 2) range 0.06% to 2.04%, 0.08% to 5.88%, and 0.14% to 17.83% for the ANN model 30-days, ANN model 10-days, and ANN model 1-day, respectively. For a total of 30 days prediction, the percentage errors for ANN model 30-days, ANN model 10-days, and ANN model 1-day are 0.25%, 1.67%, and 2.54%, respectively. This analysis assumes that the whole day or 30 days of data are not available, and the ANN model 30-days can predict the data with an error of less than 3%. For a practical case when only a small portion of data is lost, this ANN model should be able to recover the lost data with higher accuracy.
4. Conclusion
In this research, three ANN analysis models, ANN model 1-day, ANN model 10-days, and ANN model 30-days, were developed to predict the unavailable GHI data a year ago in the same month. The parameters used to train the ANN models are 1 day, 10 days, and 30 days datasets of PV output power, time, and GHI measured in July 2020. The ANN model 30-days has the most accurate prediction
where it shows the lowest RMSE and daily percentage error when comparing the predicted data with the original data in July 2019. The RMSE can be improved by increasing numbers of data during the development of an ANN model. The unavailable GHI data can be recovered with acceptable accuracy and the daily percentage error is less than 3%. The results prove that an ANN model can be used to predict the daily unavailable GHI data for performance assessment of a project or for determining the appropriate design parameters of a new renewable energy project in the same region as an industrial application.

5. References
[1] A. K. Yadav and S. S. Chandel, “Solar radiation prediction using artificial neural network techniques: A review,” *Renew. Sustain. Energy Rev.*, vol. 33, pp. 772–781, 2014.
[2] K. Y. Lai and B. H. Lim, “Optimal inverter sizing ratio for photovoltaic power plants in Malaysia,” *Proc. 11th Int. Conf. Appl. Energy*, vol. 5, part 4, ID: #774, pp. 1–5, 2019.
[3] C. Chen, S. Duan, T. Cai, and B. Liu, “Online 24-h solar power forecasting based on weather type classification using artificial neural network,” *Sol. Energy*, vol. 85, no. 11, pp. 2856–70, 2011.
[4] W. I. Hameed *et al.*, “Prediction of solar irradiance based on artificial neural networks,” *Inventions*, vol. 4, no. 3, pp. 1–10, 2019.
[5] E. A. Ahmed and M. E.-N. Adam, “Estimate of global solar radiation by using artificial neural network in Qena, Upper Egypt,” *J. Clean Energy Technol.*, vol. 1, no. 2, pp. 148–150, 2013.
[6] O. Solmaz and M. Ozgoren, “Prediction of hourly solar radiation in six provinces in Turkey by artificial neural networks,” *J. Energy Eng.*, vol. 138, no. 4, pp. 194–204, 2012.
[7] N. Premalatha and A. Valan Arasu, “Prediction of solar radiation for solar systems by using ANN models with different back propagation algorithms,” *J. Appl. Res. Technol.*, vol. 14, no. 3, pp. 206–214, 2016.
[8] D. A. Fadare, “Modelling of solar energy potential in Nigeria using an artificial neural network model,” *Appl. Energy*, vol. 86, no. 9, pp. 1410–22, 2009.
[9] M. Marzouq, Z. Bounoua, A. Mechaqrane, H. E. Fadili, Z. Lakhliai, and K. Zenkouar, “ANN-based modelling and prediction of daily global solar irradiation using commonly measured meteorological parameters,” *IOP Conf. Ser. Earth Environ. Sci.*, vol. 161, no. 1, 2018.
[10] “About PVMS – SEDA Malaysia : National PV monitoring system webportal.” https://pvms.seda.gov.my/pvportal/about/ (accessed Nov. 10, 2020).
[11] P. Truatmoraka, N. Waraporn, and D. Suphachotiwatana, “Water level prediction model using back propagation neural network,” *4th Int. Symp. Comput. Bus. Intell. ISCBI 2016*, pp. 200–205, 2016.
[12] L. Prechelt, “PROBEN1-A set of neural network benchmark problems and benchmarking rules,” Tech. rep. 21/94, *Fak. fur Inform. Univ. Karlsruhe, Ger.*, 1994.
[13] V. Bhattacherjee, “The soft computing approach to program development time estimation,” *Proc. - 9th Int. Conf. Inf. Technol. ICIT 2006*, pp. 291–292, 2007.

Acknowledgment
The research was supported by the Ministry of Higher Education Malaysia through Fundamental Research Grant Scheme (FRGS/1/2018/TK10/UTAR/02/10), Universiti Tunku Abdul Rahman through UTAR Research Fund (IPSR/RMC/UTARRF/2018-C2/04), and Malaysian Photovoltaic Industry Association’s crowdfunding (Memorandum of Agreement dated 29th July 2019). The authors would like to express sincere gratitude to the three parties for financial support. Besides, the authors would like to thank Chong Kok Keong for sharing the pyranometer for solar irradiance measurement.