Solar parabolic trough thermal energy output forecasting based on K-Nearest Neighbors approach

Jafferi Jamaludin\textsuperscript{1} and Hiromitsu Ohmori\textsuperscript{2}
\textsuperscript{1} Higher Institution’s Center of Excellence (HICoE), University of Malaya Power Energy Dedicated Advanced Center (UMPEDAC), Level 14, Wisma R & D, University of Malaya, Jalan Pantai Baharu 59990 Kuala Lumpur, Malaysia
\textsuperscript{2} Department of System Design Engineering, Faculty of Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan

E-mail: jafferi@um.edu.my

Abstract. Solar thermal output forecasting that is derived from the solar irradiance forecasting, is very much exposed to high forecasting error. This is because of the heavy dependence on the solar irradiance prediction accuracy that could be very low in certain situations owing to the high uncertainty in weather conditions. By considering this fact, this paper proposes to develop a solar thermal output forecasting model using measured solar irradiance data, instead of the predicted data. The proposed model applies K-Nearest Neighbors (K-NN) algorithm to generate 24-hour ahead forecasting data on solar thermal output from a solar parabolic trough system. For the purpose of illustrating the forecasting model performance, Kuala Lumpur, Malaysia is used as a case study, with PolyTrough 1800 model is selected as the solar parabolic trough collector under investigation. Simulation has been carried out using Matlab software to verify the effectiveness of the proposed K-NN-based forecasting model. The results show that the model is able to produce acceptable results in certain conditions.

1. Introduction
Concentrating solar power technology has rapidly grown in recent years owing to its ability to provide high temperatures in the range of 100°C to 450°C or even more. The technology basically focuses the sunlight at a receiver to achieve high temperatures. One of the most popular receivers is the solar parabolic trough concentrator. This concentrator offers several advantages which include high power density, high efficiency, modularity and versatility [1]. In addition, since it focuses on the direct radiation from the sun, a tracking system is required. The fact that this concentrator can reach high temperature range makes the solar parabolic trough system popular as a source of electricity since higher receiver temperature leads to higher efficiency and thus, results in better electricity generation.

Despite the strengths of the solar parabolic trough system, its solar thermal output is very much dependent on weather conditions. This introduces uncertainty in electricity generation which consequently causes power fluctuation, voltage stability risk and energy imbalance between supply and demand. As a means to reduce this uncertainty, solar irradiance forecasting has been widely employed [2]. The forecasting result is then used to estimate the solar thermal output. A number of studies have been conducted to forecast solar thermal output based on solar irradiance forecasting results. In [3], two Numerical Weather Prediction (NWP) models namely the Aerosol-based Forecasts of Solar Irradiance for Energy Applications (AFSOL) and the European Centre for Medium-range
Weather Forecasts (ECMWF) in combination with the Skartveit and Olseth global-to-diffuse model have been investigated to forecast the direct normal irradiance. The forecasting results are then converted to solar thermal output prediction based on the physical model of the parabolic trough solar thermal plant. In another study [4], by using the parabolic trough model as well, the solar thermal output forecasting is derived from the direct normal irradiance forecasting using the ECMWF together with the High-Resolution Limited Area Model (HIRLAM). [5] proposes the forecast technique of a one-day-ahead hourly thermal energy collection based on solar radiation forecasting by using three neural network models: feedforward neural network, radial basis function neural network and recurrent neural network. A model that combines the meteorological forecasts and thermal performances of a solar Fresnel power plant is presented in [6]. By feeding the model with the forecasted direct normal irradiance which is inferred from the global horizontal irradiance, the power output of the plant can be estimated.

Most of the studies found in the literature translate the solar thermal output forecasting from the solar irradiance forecasting. This approach is useful for short-term forecasting whereby information on the solar thermal output is needed a few minutes to a few hours ahead. However, forecasting of this type depends very much on the accuracy of the solar irradiance forecasting. On the other hand, for cases where the measured solar irradiance data is available from the previous days, forecasting of the solar thermal output can be based on the measured data, instead of the predicted data. Hence, the forecasting error of the solar thermal output can be reduced since the dependence on the solar irradiance forecasting accuracy has been removed. Taking this into account, this paper presents the 24-hour ahead forecasting model of the solar thermal output from a parabolic trough system based on the available solar irradiance data. The K-Nearest Neighbors (K-NN) technique is proposed to develop the forecasting model. Section 2 details the proposed K-NN-based forecasting model. Section 3 provides a description of the case study used to verify the performance of the proposed model. Section 4 presents the simulation results while Section 5 draws the conclusion.

2. Forecasting model

K-NN is attractive because of the simplicity and ease of implementation. Moreover, in certain circumstances, it outperforms the more complex algorithms in producing results with better accuracy. In this study, K-NN is used to forecast solar thermal output 24 hours ahead based on the assumption that unknown observations which have high similarity to known observations should belong to the same class or should produce almost the same outcomes.

For the proposed K-NN model, a database of solar thermal output with n observed data points is constructed based on the measured solar irradiance data obtained. Determination of the solar thermal output is done by first calculating the solar thermal collector efficiency using (1) below [7]:

\[
\eta = \eta_0 - a_1 \frac{T_{in} - T_{am}}{G} - a_2 \left( \frac{T_{in} - T_{am}}{G} \right)^2
\]  

(1)

Here, \( \eta \) is the collector efficiency, \( \eta_0 \) is the maximum collector efficiency, \( a_1 \) is linear heat loss coefficient in Wm\(^{-2}\)K\(^{-1}\), \( a_2 \) is quadratic heat loss coefficient in Wm\(^{-2}\)K\(^2\), \( T_{in} \) is the inlet temperature in K, \( T_{am} \) is the ambient temperature in K and \( G \) is the solar beam irradiance in Wm\(^{-2}\). By using the calculated collector efficiency, the solar collector output \( S_j \) is obtained using (2):

\[
S_j = \eta_j G_j \quad \text{with} \quad j = 1, 2, 3 \ldots, n
\]  

(2)

For the 24-hour ahead forecasting to take place, the previous 24-hour solar collector output data series is taken as the base data points. In order to assess the similarity between an observed data point in the database of the solar thermal output and a base data point, the distance metric between the two data points is calculated. In this study, Euclidean distance as given in (3) is used:

\[
d(S_b, S_j) = \left[ (S_b - S_j)^2 + (t_b - t_j)^2 \right]^{1/2}
\]  

(3)
\( S_b \) is the base data point while \( t_b \) and \( t_j \) are the times when \( S_b \) and \( S_j \) occur respectively. By using (3), the distances between all base data points and all observed data points are determined. For each base data point, all calculated distances are sorted in ascending order so that the observed data points are ranked with the nearest points at the top of the order. These nearest points which are usually referred to as the nearest neighbors, are then assigned with different weights. The closer the nearest neighbors to the base data point, the higher the weight is. This results in closer nearest points to contribute more to the forecasting. Here, Gaussian weighting function is used as expressed below:

\[
w(S_b, S_j) = \exp \left( -\frac{d(S_bS_j)^2}{\sigma^2} \right)
\]  

(4)

\( \sigma \) is the width of the Gaussian function which needs to be properly chosen. In order to calculate the output that corresponds to each base data point, (5) applies:

\[
S_o = \frac{\sum_{j=1}^{N} w(S_bS_j)S_j}{\sum_{j=1}^{N} w(S_bS_j)}
\]  

(5)

\( S_o \) is taken as the forecasted solar thermal output which depends on the number of nearest neighbors considered as denoted by \( k \) in (5). It is a normal practice to select \( k \) that leads to the minimum forecasting error. As a means to estimate the forecasting error, it is commonly accepted to use percent mean absolute error (%MAE), mean absolute error (MAE) and root mean square error (RMSE) metrics as given below:

\[
%MAE = \frac{\sum_{i=1}^{N} |S_{m,i} - S_{o,i}|}{\sum_{i=1}^{N} S_{m,i}} \times 100\% \quad (6)
\]

\[
MAE = \frac{1}{N} \sum_{i=1}^{N} |S_{m,i} - S_{o,i}| \quad (7)
\]

\[
RMSE = \left[ \frac{1}{N} \sum_{i=1}^{N} (S_{m,i} - S_{o,i})^2 \right]^{\frac{1}{2}} \quad (8)
\]

In (6)–(8), \( S_{m,i} \) and \( S_{o,i} \) are respectively the measured and forecasted solar thermal outputs at the \( i \)th hourly interval.

3. Case study

For the purpose of testing the forecasting model described in Section 2, a case study has been conducted. In this study, Kuala Lumpur which is the capital of Malaysia has been chosen. In order to execute the forecasting algorithm, solar thermal output from a solar parabolic trough system is estimated. In this study, the parabolic trough collector of model PolyTrough 1800 from NEP Solar AG [8] is used. Based on (1), the following equation is obtained.

\[
\eta = 0.689 - 0.36 \frac{t_{in} - t_{am}}{G} - 0.0011 \left( \frac{t_{in} - t_{am}}{G} \right)^2
\]  

(9)

where \( \eta_0 = 0.689 \), \( a_1 = 0.36 \text{ W/K}^{-1}\text{ m}^{-2} \) and \( a_2 = 0.0011 \text{ W/K}^{-2}\text{ m}^{-2} \) [8]. By using (9), the solar collector efficiency can be calculated. From the calculated efficiency, the solar thermal output can be achieved by using (2). Here, the solar beam irradiation can be obtained from any source that provides available measured solar irradiation data. In this study, the data is taken from the photovoltaic geographical information system (PVGIS) [9]. For the purpose of illustrating the performance of the forecasting algorithm, the PVGIS data is assumed to be of measured values.

4. Simulation results
Simulation has been carried out using Matlab software to evaluate the effectiveness of the proposed forecasting model. The solar irradiance data taken from the PVGIS are between January 2016 and April 2016 for the case of Kuala Lumpur, Malaysia. For the purpose of calculating the solar collector efficiency using (9), it is assumed that $T_{in}$ is 150°C and $T_{am}$ is obtained from the weather data provided by PVGIS during the aforementioned period. The database of the observed data points comprises the solar thermal output data from 1 January 2016 until 7 April 2016 which are calculated using (2).

As for the calculated solar thermal output data between 8 April 2016 and 14 April 2016, they are treated as the base data points to derive the 24-hour forecasting data on 9, 10, 11, 12, 13, 14 and 15 April 2016. Here, forecasting on 9 April uses the 24-hour calculated solar thermal output data on 8 April as the base data points. The same goes for the 10 April forecasting which utilizes the 9 April 24-hour calculated solar thermal output data as the base data points. This is the same for other days as well.

The simulation results are shown in figure 1 considering $k$ equals 20 and $\sigma$ equals 0.5. It can be seen in figure 1(a) that forecasting on 9 April and 10 April record significant errors which are 103% and 47% respectively in terms of the %MAE. This happens as a result of a considerable difference in terms of the amount and hourly pattern of solar thermal output between 8 April and 9 April as well as between 9 April and 10 April. Since the forecasting uses the previous 24-hour data from the day before as the base data points, the forecasting algorithm is not able to produce the desired prediction results. As a means to address this shortcoming, the previous 1-week data is used to generate the average 24-hour data to be utilized as the base data points. As an example, in order to carry out the forecasting for 9 April, the base data points take the average hourly data from 2 April to 8 April, instead of 8 April only. Figure 1(b) displays the simulation results. It can be observed that the %MAE for 9 April and 10 April have reduced to 81% and 25% respectively. Figure 2, figure 3 and figure 4 details the error comparison. Despite the error reduction for 9 April and 10 April, there is however an increase in the errors for 11 April and 12 April. Nonetheless, in terms of the overall performance, the previous 1-week average approach records lower overall average errors for the 7 days as presented in Table 1.

![Figure 1](image1.png)

**Figure 1.** Comparison between measured and predicted data of solar thermal output based on: (a) the previous 1-day data approach and (b) the previous 1-week average data approach.

![Figure 2](image2.png)

**Figure 2.** Comparison of %MAE.

![Figure 3](image3.png)

**Figure 3.** Comparison of MAE.
5. Conclusion

In this paper, K-NN approach has been proposed for 24-hour ahead solar thermal output forecasting from a solar parabolic trough system using measured data, instead of predicted data derived from solar irradiance forecasting. Two approaches to generate the base data points for the K-NN algorithm have been presented: previous 1-day and the previous 1-week average approaches. It has been demonstrated that the previous 1-week average approach produces forecasting results with lower errors in most instances. In addition, the previous 1-week average approach not only reduces the errors but also flattens the error curves. In conclusion, K-NN approach could be considered as a reliable approach to forecast solar thermal output without depending on the solar irradiance forecasting results. For future work, additional inputs influencing solar thermal outputs such as sun elevation, wind speed and rainfall will be considered to further improve the forecasting results.

6. References

[1] Hafez A Z, Attia A M, Eltawab H S, ElKousy A O, Afifi A A, AbdElhamid A G, AbdElkader A N, Fateen S-E K, El-Metwally K A, Soliman A and Ismail I M 2018 Design analysis of solar parabolic trough thermal collectors Renew. Sustain. Energy Rev. 82 1215–60

[2] Law E W, Prasad A A, Kay M and Taylor R A 2014 Direct normal irradiance forecasting and its application to concentrated solar thermal output forecasting – a review Sol. Energy 108 287–307

[3] Wittmann M, Breitkreuz H, Schroedter-Homscheidt M and Eck M 2008 Case studies on the use of solar irradiance forecasts for optimised operation strategies of solar thermal power plants IEEE J. Sel. Topics Appl. Earth Obs. Remote Sensing 1 18–27

[4] Kraas B, Schroedter-Homscheidt M and Madlener R 2013 Economic merits of a state-of-the-art concentrating solar power forecasting system for participation in the Spanish electricity market Sol. Energy 93 244–255

[5] Yona A and Senjyu T 2012 Thermal energy collection forecasting based on soft computing techniques for solar heat energy utilization system Smart Grid Renew. Energy 3 214–21

[6] Rodat S, Tantolin C, Le Pivert X and Lespinats S 2016 Daily forecast of solar thermal energy production for heat storage management J. Clean. Prod. 139 86–98

[7] Tzivanidis C, Bellos E, Korres D, Antonopoulos K A and Mitsopoulos G 2015 Thermal and optical efficiency investigation of a parabolic trough collector Case Stud. Therm. Eng. 6 226–37

[8] Solar collector factsheet NEP Poly Trough 1800

[9] Huld T, Müller R and Gambardella A 2012 A new solar radiation database for estimating PV performance in Europe and Africa Sol. Energy 86 1803-1815

Acknowledgments

This paper was produced as part of the research support program sponsored by The Hitachi Global Foundation and JST CREST Grant Number JPMJCR15K5 (Hayashi Team), Japan. The authors would like to express their appreciation to the sponsors for the support given.