Development of Near-Real-Time Solar Generation Prediction Technique Using Weather Data

Navareen Sohkhlet and Bikramjit Goswami

Abstract This paper presents a technique for forecasting solar power generation using weather forecast data. Solar power generation mainly depends on the relative position of the sun and some extrinsic as well as intrinsic factors. Extrinsic factors, such as cloud cover, temperature, rainfall, humidity, and wind speed, are used for the prediction of solar generation. Apart from these, the intrinsic factors are also taken as inputs for the proposed prediction technique. The artificial intelligence–based techniques of linear regression, polynomial regression, and artificial neural networks are used for prediction purposes, with input data of all the months of the year 2021. After developing different AI models, their accuracies are compared before selecting the best technique for solar generation prediction. The AI model found to be accurate during the present work is applicable to all solar generation systems, for generation level prediction.

Keywords Solar generation prediction · Linear regression · Polynomial regression

1 Introduction

Solar power is one of the most promising renewable energy sources, the generation of which does not result in the emission of pollutants and greenhouse gases (Kim et al. 2017). Global warming and the energy crisis over the past few decades have motivated the use and development of alternative, sustainable, and clean energy sources (Sabzehgar et al. 2020). Solar energy or solar power is a nonconventional source or renewable source of energy and its intensity depends on the weather and sun's position concerning the panels. Electricity demand needs to be predicted for better planning of utilization. Solar energy generation varies frequently with the weather and relative solar position (Verma et al. 2016).

N. Sohkhlet · B. Goswami (✉)
Assam Don Bosco University, Assam, India
e-mail: bikramjit.goswami@dbuniversity.ac.in

Solar power forecasting is a complex process as it mainly depends on climate conditions, which change or fluctuate over time. Therefore, machine learning (ML)–based methods have been used for effective solar power generation forecasting. For improved accuracy, new and more intelligent methods are being developed (AlKandari and Ahmad 2020).

Renewable energy or nonconventional generation resources such as solar and wind behave in a stochastic fashion due to frequent weather parameter changes. Transmission and distribution losses also affect the performance of these resources. Therefore, scheduling, optimization, and management of smart grids and micro-grids with a high concentration of renewable energy resources are the main issues or challenges of such grids. One of the most promising practices in scheduling the performance of smart grids is to forecast the energy production of the resources that lead to energy generation and cost-efficient replacement of the current process (Sabzehgar et al. 2020). Models based on artificial intelligence (AI) are used for optimization and control purposes due to their learning capabilities. Examples of these models include support vector machines (SVM), regression, and neural networks.

AI techniques play an important role in prediction modeling and analysis of the performance as well as in controlling renewable energy generation processes. AI techniques are used to solve complicated problems and practices in various fields of the engineering and technology. AI systems can be used as a way to solve complex problems. AI systems have been used in diverse applications of pattern recognition, manufacturing, optimization, control, robotics forecasting, power systems, signal processing, medical, and social sciences (Kumar and Kalavathi 2018). Their use in renewable energy generation forecasting is however a newer application, becoming increasingly popular nowadays.

2 Use of AI in Solar Generation Prediction

One of the simplest AI-based techniques used in generation prediction is regression analysis, which is effectively used for solar generation prediction in the present work. It is a method to determine a functional relationship of the model between predictor (independent) parameters and response (dependent). The regression analysis shows a repetitive process so that the outputs can be utilized to verify, criticize, analyze, and modify the inputs (Nalina et al. 2014). The regression approach identifies the correlation between parameters by fitting the linear equation to the data. Figure 1 shows the plot between solar generation and ambient temperature. This type of distribution of data can be analyzed using the regression technique very conveniently.

The methodology of modeling the system for the prediction of solar generation is shown in Fig. 2. It shows the stepwise procedure of the AI model developed for obtaining the forecast data considering weather conditions, such as, sunny, scattered cloud, partly sunny drizzle over, overcast, thunderstorm, etc. Another input is daily solar generation in net million units (MU) from the North Eastern Regional Load Dispatch Centre (NERLDC) for the state of Assam for the year 2021.
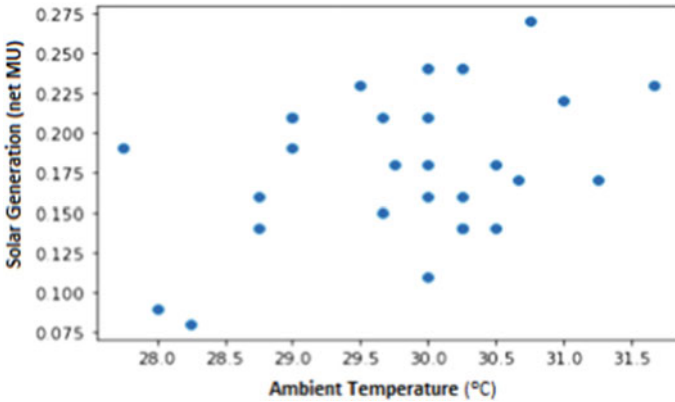
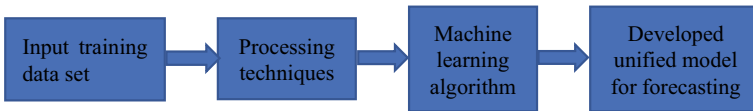**Fig. 1** The plot of solar generation versus ambient temperature

Fig. 2 Block diagram of the system model

Figure 3 shows the step-by-step methodology of training, testing, and validation of a regression-based AI model designed for predicting solar generation. The weather parameters are gathered from different authentic weather forecast websites. Initially, the solar generation is predicted using a model generated, based on the aforementioned weather parameters, using two different approaches: linear and polynomial regression techniques. Then, these estimated values are used to forecast the generated power in the grid. The estimation models are obtained from the averaged data gathered in the year 2021. An essential step in obtaining forecast models is to have two sets of data: the training and the test data sets. The training data set is used to obtain the prediction model, and the test data set is used for model evaluation.

Therefore, further weather and generation data were gathered to test the accuracy of the model. The input data are all sorted before being used. The data with null entry are eliminated, and the data are normalized to increase the accuracy of the forecast models. After utilizing these models for forecasting solar power generation, the error of each model is measured by comparing the original target data and the forecast data. The analysis and predictions are performed using Python software for programming and training regression algorithms.

As mentioned earlier, the solar energy generation forecast, utilized in this work, is done in phases. Phase one is the data preprocessing for the forecast models, which in the case of the current work includes eliminating all of the null entry data. In phase two, a model for estimating and forecasting solar generation based on weather parameters is derived. It can be comprehended from these figures that while some of
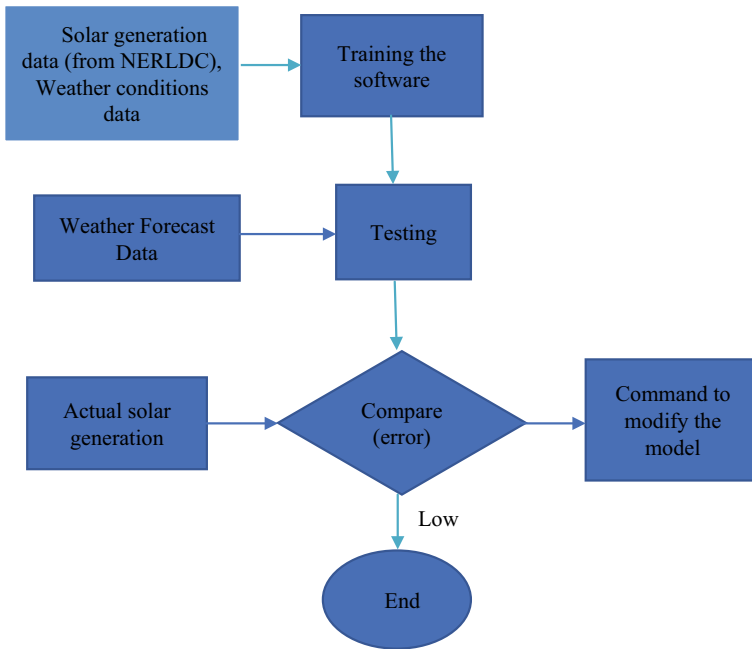
Fig. 3 Process flowchart of the AI model

the weather parameters are linearly related to solar generation, some are not. Hence, polynomial regression technique is found to be more accurate in estimating solar generation as compared to linear regression.

3   Results and Discussion

Figure 4 shows the best fit polynomial graph after the regression technique is implemented for solar generation estimation in August 2021. In the figure, the horizontal axis shows the ambient temperature, and the vertical axis shows solar generation data obtained from NERLDC for the state of Assam.

Table 1 shows the estimated solar generation considering 1-month-long data for training the regression model and predicting the solar generation of the subsequent days for the different seasons of the year 2021.

Table 1 shows the accuracy of estimated solar generation for different days of the seasons of 2021 for training and validation, using polynomial regression, thereby considering the actual solar generation (net MU) from the NERLDC website and the predicted solar generation (net MU) from the technique developed and subsequently computing the relative error. Figure 5 shows the error percentage curve by taking the actual and the predicted value obtained from the system forecasting model using
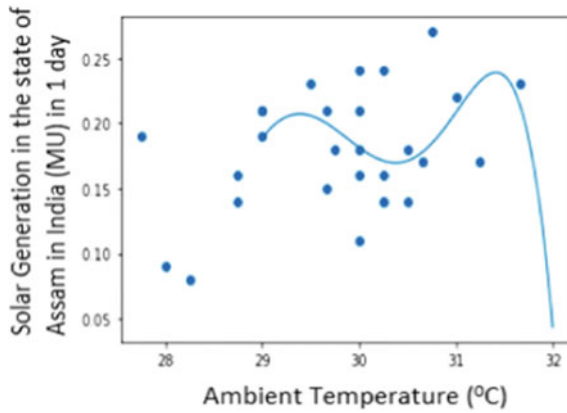
Fig. 4 Polynomial fit of solar generation vs. ambient temperature

Table 1 Estimated solar generation using polynomial regression using month-long data for different seasons in the year 2021

Training period	Estimation date	Actual solar generation (net MU)	Predicted solar generation (net MU)	Error (%)
1–28 February	1 March	0.17	0.1479	13
1–31 May	1 June	0.15	0.129	14
1–31 August	1 September	0.15	0.182	−21
1–30 September	1 October	0.29	0.23	20
1–31 October	1 November	0.20	0.21	−5

Fig. 5 Error percentage curve between the actual and the predicted solar generation in (net MU)

polynomial regression technique. The error percentage lies within the range of -
5% to +20% for the period considered. Table 2 shows estimated solar generation
considering data for 15 days of the year 2021 for training. The methodology used
is the same as the training done using 1-month data. The accuracy calculations are
also shown for the different seasonal estimations for the year. The table shows the
calculation of the accuracy of different months of 2021 for training and validation
using the regression algorithm, thereby considering the actual solar generation (net
MU) from the NERLDC website and the predicted solar generation (net MU) from
the trained regression model and obtaining the error.

Figure 6 shows the error curve by taking the actual and the predicted value obtained
from the system forecasting model using polynomial regression technique for a
particular period interval as shown in Table 2 and its error percentage lies within the
range of −23% to +17%.

Table 3 shows estimated solar generation considering weekly data (7 days each)
for the training of the prediction system for different seasons of the year 2021. In
the table, the accuracy of the estimation technique considering weekly data as input
(weather conditions, temperature, solar generation, etc.) for different seasons of the
year 2021 for training. By comparing the actual and predicted solar generation, it can
be observed that training the system with weekly data gives better accuracy. With
month-long data for training, the error is slightly more than that with week-long data.
With 15 days of data used for training, the error in estimating solar generation is the
highest. Training the regression model with less than 7 days data leads to greater
error. Hence, it is observed that the optimum period of training of the regression
model for a more accurate prediction of solar generation is 7 days (1 week).Fig. 7
shows the error curve by taking the actual and the predicted values obtained from the
system forecasting model using polynomial regression techniques for a particular
period of interval as shown in Table 3, and its error percentage lies within the range
of −25.5% to +15%.

Table 2 Estimated solar generation using polynomial regression using 15 days of data for different
seasons in the year 2021

Training period	Estimation date	Actual solar generation (net MU)	Predicted solar generation (net MU)	Error (%)
1–15 January	16 January	0.06	0.089	−4.8
1–14 February	15 February	0.1	0.1236	−23
15–28 February	1 March	0.17	0.14	17
1–15 March	16 April	0.14	0.149	−6.4
1–15 May	16 May	0.02	0.018	10
1–15 July	16 July	0.19	0.1939	−2
16–31 July	1 August	0.19	0.1384	16
15–30 September	1 October	0.29	0.296	−0.2
1–15 October	16 October	0.24	0.27	−10
1–15 November	16 November	0.26	0.25	3

Fig. 6 Error percentage curve between the actual and the predicted solar generation in (net MU) using 15-day data period interval

Table 3 Estimated solar generation using polynomial regression using 7 days of data for different seasons in the year 2021

Training period	Estimation date	Actual solar generation (net MU)	Predicted solar generation (net MU)	Error (%)
1–6 October	7 October	0.26	0.28	−10
7–13 October	14 October	0.29	0.295	−1.79
21–26 October	27 October	0.26	0.221	15
1–6 November	7 November	0.29	0.2922	−0.75
14–20 November	21 November	0.24	0.236	1.6
21–26 November	27 November	0.26	0.226	13
1–6 December	7 December	0.2	0.255	−25.5
14–20 December	21 December	0.22	0.253	−15
21–26 December	27 December	0.24	0.243	−1.25

Fig. 7 Error percentage curve between the actual and the predicted solar generation in (net Mu) using 7-day data period interval

4 Conclusion

In this work, an AI-based solar generation prediction model is developed, based on the actual values of different weather parameters for the city of Guwahati, Assam, in the year 2021. The designed regression-based AI model gives accuracy in the range of 0.75–25.5% error. The model can be further improved in accuracy by incorporating more inputs, such as near-real-time remote sensed data. Moreover, as the inputs are highly nonlinear in relation, the use of artificial neural networks in the training is expected to provide more accuracy in estimating solar generation in the region. This AI-based technique also requires testing and validation for other areas of the world for the global validity of the model.

References

AlKandari M, Ahmad I (2020) Solar power generation forecasting using ensemble approach based on deep learning and statistical methods. Appl Comput Inf

Kim JG, Kim DH, Yoo WS, Lee JY, Kim YB (2017) Daily prediction of solar power generation based on weather forecast information in Korea. IET Renew Power Gener 11(10):1268–1273

Kumar KR, Kalavathi MS (2018) Artificial intelligence based forecast models for predicting solar power generation. Mater Today Proc 5(1):796–802

Nalina U, Prema V, Smitha K, Rao KU (2014) Multivariate regression for prediction of solar irradiance. In: International conference on data science and engineering (ICDSE). IEEE, pp 177–181

Sabzehgar R, Amirhosseini DZ, Rasouli M (2020) Solar power forecast for a residential smart microgrid based on numerical weather predictions using artificial intelligence methods. J Build Eng 32:101629

Verma T, Tiwana APS, Reddy CC, Arora V, Devanand P (2016) Data analysis to generate models based on neural network and regression for solar power generation forecasting. In: 7th international conference on intelligent systems, modelling and simulation (ISMS). IEEE, pp 97–100