2024 32nd Southern African Universities Power Engineering Conference (SAUPEC) | 979-8-3503-7134-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/SAUPEC60914.2024.10445076

Short-Term Hours Ahead Forecast of Expected Available Solar Power Using Linear Regression Machine Learning Scheme

Adedayo Ademola Yusuff, Thapelo Cornelius Mosetlhe
Department of Electrical Engineering
University of South Africa
Johannesburg, South Africa
Email: yusufaa@unisa.ac.za, mosettc@unisa.ac.za

***Abstract*—Integration of renewable energy sources which are intermittent in nature can exacerbate the inherent challenges in hours ahead operational planning. In this work a machine learning scheme based on linear regression is used to forecast hours ahead generation using a year data that is partitioned into daily and weekly categories. The performance measures of L_2 norm of error, Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) are used to assess the scheme. The results show that forecast error based on RMS, of 1 hour ahead is lower (RMS = 0.19717) than that of 3 hours ahead (RMS = 0.86876) for all categories of data partitions. In addition, the performance errors are lower when weekly partition (RMS = 0.05575) is used compared to a daily partition (RMS = 0.19717) is used in making a forecast for 1 hours ahead forecast.**

***Index Terms*—One-hour ahead forecast, two hours ahead forecast, solar power, renewable energy sources, intermittent sources, linear regression, machine learning, expected power estimation, mean absolute error, root-mean-square error.**

I. Introduction

In electric power system planning and operation, mid-to-long-term load and energy forecast of power is essential. Unlike the traditional fossil fuel power generating stations, power from renewable energy sources such as wind and solar is heavily reliant on climatic conditions. Forecast of future climatic conditions and short-to-mid-term renewable energy availability using machine learning can improve operational efficiency. In the past, load demand has always been laden with certain level of uncertainty compared to power generation. The uncertainty in generation of power as a result of generator failure is usually catered for by spinning-reserved. Spinning reserve is wasteful, nevertheless, it is essential. Integration of renewable energy sources that are intermittent will introduce additional uncertainties to power generation.

The stochastic nature of solar energy poses a major challenges in using machine learning to predict expected available wind energy. Like all modelling and predictive schemes, machine learning requires high-quality data, salient features and appropriate algorithms. Inclusion of intermittent energy sources can exacerbate the inherent challenges in generation constrained electric power systems. It is important to have models of expected energy source availability in the future for planning and operational purpose. In view of this, the use of machine learning and other computational schemes are crucial for predicting expected solar energy.

Machine learning algorithms such as decision trees [1, 2], linear regression [3, 4], support vector machines [5], and random forests [6–8], have been used in predictive models and data science. Few of the machine learning techniques that have been used in short-time forecast of solar power generation are attention-based Bayesian Seq2Seq scheme [9]; artificial neural network (ANN), Recurrent Neural Network (RNN) and Convolutional Neural Network-Long-Short Term Memory (CNN-LSTM) [10]; univariate and multivariate models [11]; machine learning with pearson correlation [12]; XGBoost, LightGBM and CatBoost algorithms [13]; kernel ridge regression (KRR) [14]. In [9], decomposition-ensemble framework that uses ensemble empirical mode decomposition with independent component analysis and adaptive noise was used to extract the intrinsic modes of the solar power generation time series. This was used with combination of an attention-based Bayesian sequence-to-sequence (Seq2Seq) to forecast a short-term solar power generation. Apart from using machine learning schemes to deal with planning and operational issues of expected available power forecast, machine learning can also be used to improve short-term energy forecast accuracy for a given site or an unknown geographical area before installation. Deep learning algorithms based on Long Short-Term Memory (LSTM), Gated Reference Unit (GRU) and Recurrent Neural Network (RNN) were proposed in [15].

Although various methods have been used to forecast hours ahead available solar power, however, to the best of the author's knowledge, the effect of categories of data used in hours ahead forecast have not been addressed. In this work a machine learning scheme based on linear regression is used for hours ahead forecast based on daily and weekly categories of data.

The outline of this paper is as follows. In section II we present the methodology for data collection and computer experiments. Results and discussion are presented in section II-B,

This work is based on research supported wholly by National Research Foundation South Africa (Grant Numbers **150572**)

while the conclusion is given in section IV.

II. Methods

A. Data collection

The site used in the study is University of South Africa (Unisa), Science campus, Florida, South Africa. The site is located at latitude -26.158 and longitude 27.901. Historical data of year 2022 for weather and expected solar power generation at hour interval for the site was collected from Ninja [16]. The expected solar power generation from the site is based on an installed capacity of 30 kW. The historical data collected are wind speed (m/s^2), Air temperature (°C), Precipitation (mm / hour), Air density (kg / m^3), irradiance diffuse (W / m^2), irradiance direct (W/m^2), Ground-level solar irradiance (W / m^3), Top of atmosphere solar irradiance (W / m^2), Cloud cover fraction. A total of ten features were collected. The data was subsequently processed using DataFrame.jl a Julia programming language package [17].

The distribution of wind power generation at the site is shown in Fig. 1, while scatter plots of some features are shown in Fig. 2 and Fig. 3. In the distribution of electricity generated from solar power shown in Fig. 1, it is evident that a large portion of electricity generation is below the site installed capacity of 30 kW. Fig. 2 depicts the scatter plot of solar power against wind speed.

Solar power is partitioned into hourly period from 00H00 to 23H00. The scatter plot of solar power and hourly period catigorisation is shown in Fig. 4. It is evident from Fig. 4 that electricity generated from solar power varies from 0 kW to 30 kW.

B. Modelling and Simulation

The data obtained from the investigation site consist of 8760 observations. The data was partitioned into two categories, namely daily and weekly categories. The daily and weekly categories consist of 365 and 52 folds respectively. Each of the folds is subdivided into training and test set depending on number of hours ahead. Assuming a fold consist of N observations, and h is hour ahead. The number of observation N_{tr} for a training set and test set N_{ts}, are $N_{tr} = N - h$ and $N_ts = h$ respectively.

A linear regression machine learning scheme was designed in a Julia programming language environment [17, 18]. Models were formulated based on categories of data, number of folds and the afore mentioned machine learning scheme. Each model was then subsequently simulated on a high performance computing (HPC). The HPC has 128 compute nodes which are dual socket servers with 14 cores per socket, resulting in 28 cores per node, for a total of 3584 cores across the platform. The nodes use Intel E5-2690 v4 CPU's at 2.60GHz. Each of the nodes has 256GB RAM.

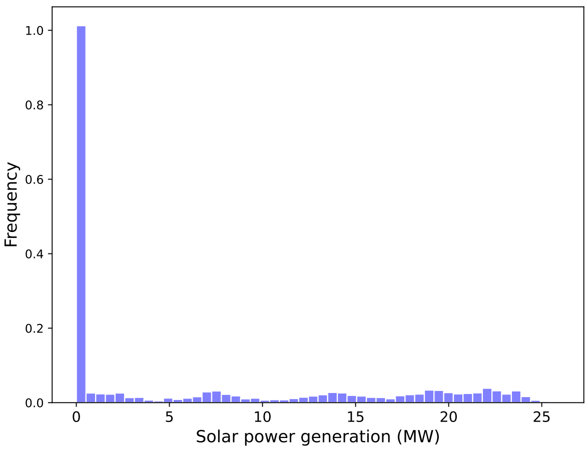

Fig. 1: Solar power generation

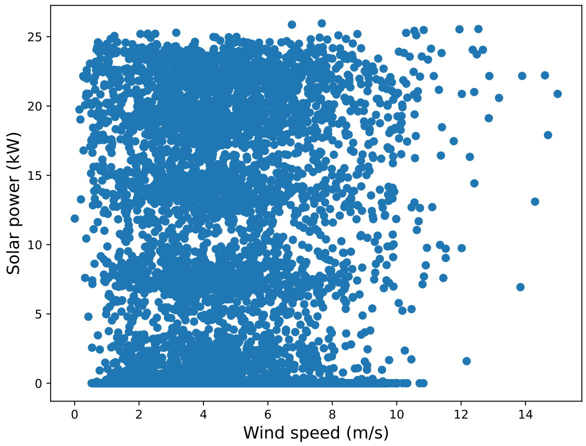

Fig. 2: Scatter solar generation vs wind speed

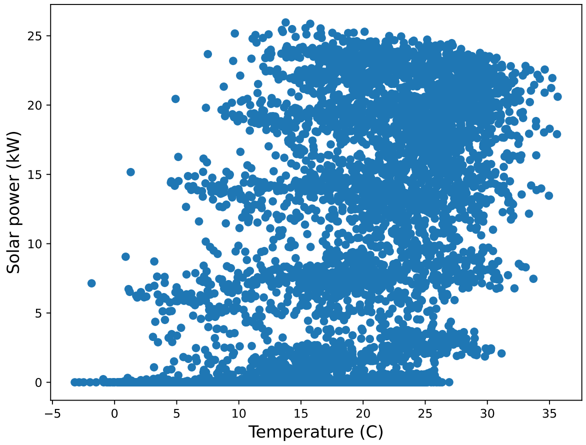

Fig. 3: Scatter plot of solar generation vs temperature.

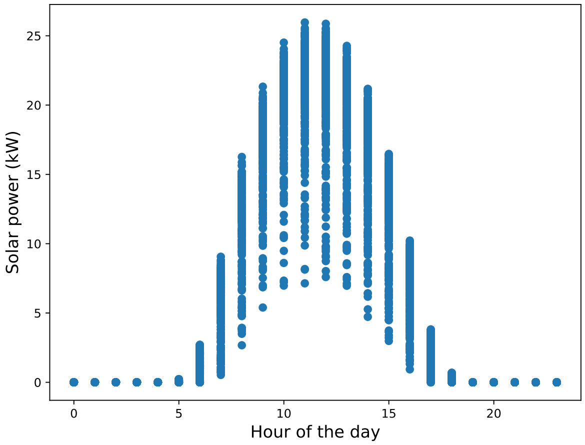

Fig. 4: Scatter plot of solar generation vs period of the day.

III. Results and Discussion

The performance error of the machine learning scheme for categories of time ahead forecast and partitions are presented in Table I. It is evident in the table that the result for one hour ahead based on a partition of a week gives the best performance.

TABLE I: Performance of linear regression based machine learning scheme for hours a head forecast.

hours a head	Partions	L_2	MAE	RMS
1.0	365	0.03888	0.14165	0.19717
1.0	52	0.00311	0.0453	0.05575
3.0	365	0.75475	0.4992	0.86876
3.0	52	0.00645	0.05683	0.08031

The comparison of specific categories of hours ahead and initial lag training time are presented in Fig. 5 to Fig. 7. In all the three figures, the forecast errors for one-hour ahead forecast are lower than the ones for three hours ahead forecast.

Fig. 8 depicts the plots of the forecast rms errors of one and three hours ahead based on one sun day. It should be noted that the initial forecast errors of one and two hours for a one sun day are significantly large. However, as the number of observations that are used to train the scheme increases, the performance improves for both one hour and three hours ahead forecast.

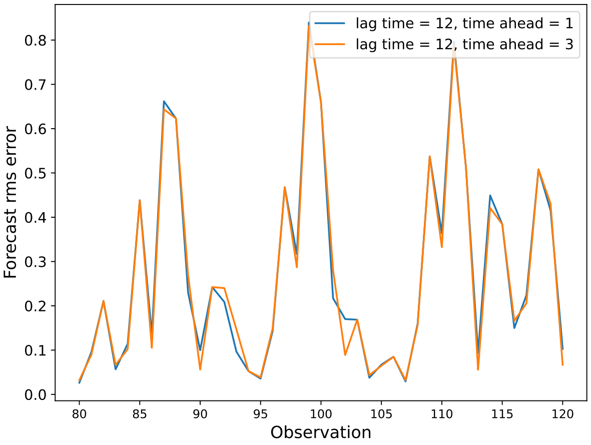

Fig. 5: Lag of one sun day, 1 and 3 hours ahead.

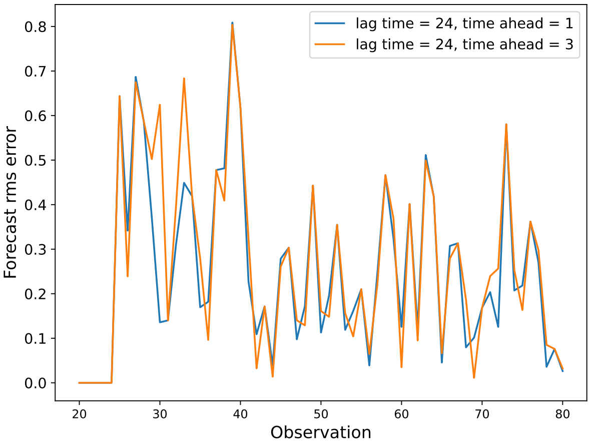

Fig. 6: Lag of two sun day, 1 and 3 hours ahead.

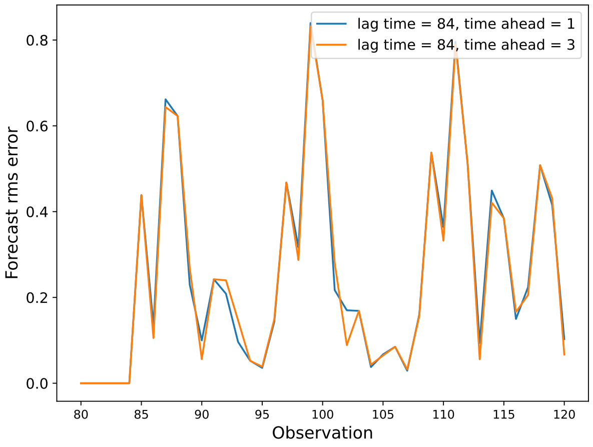

Fig. 7: Lag of seven sun day, 1 and 3 hours ahead.

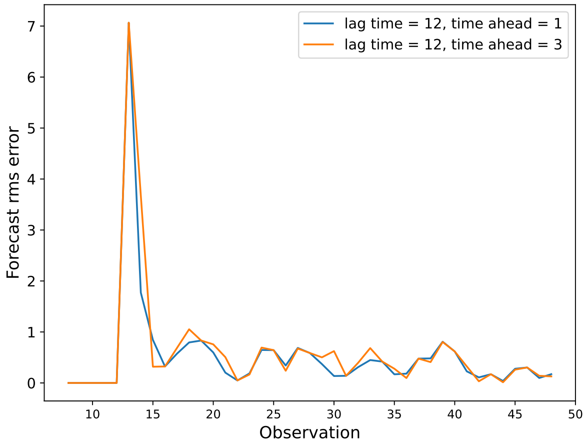

Fig. 8: Initial forecast rms error: lag of one sun day (one sun week), 1 and 3 hours ahead.

IV. Conclusion

Load demand has always been laden with certain level of uncertainty compared to power generation, consequently, integration of renewable energy sources that are intermittent at high level of penetration will introduce additional uncertainties to power generation. Hence, it is important to have a forecast scheme that can predict few hours ahead of estimated available solar power for short to medium-term operational planning. We used linear regression machine learning scheme to forecast one (1) hour and three (3) hours ahead of expected solar power at University of South Africa, Florida campus. The rationale for using a linear regression machine learning scheme is because it is computationally efficient. The results show that a linear regression machine learning scheme trained with weekly data produced a better forecast compared with the same scheme trained on daily data.

References

[1] B. De Ville, "Decision trees," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 5, no. 6, pp. 448–455, 2013.

[2] S. B. Kotsiantis, "Decision trees: a recent overview," *Artificial Intelligence Review*, vol. 39, pp. 261–283, 2013.

[3] G. K. Uyanık and N. Güler, "A study on multiple linear regression analysis," *Procedia-Social and Behavioral Sciences*, vol. 106, pp. 234–240, 2013.

[4] G. James, D. Witten, T. Hastie, R. Tibshirani, and J. Taylor, "Linear regression," in *An Introduction to Statistical Learning: With Applications in Python*. Springer, 2023, pp. 69–134.

[5] F. Zhang and L. J. O'Donnell, "Support vector regression," in *Machine learning*. Elsevier, 2020, pp. 123–140.

[6] X. Zhou, X. Zhu, Z. Dong, W. Guo *et al.*, "Estimation of biomass in wheat using random forest regression algorithm and remote sensing data," *The Crop Journal*, vol. 4, no. 3, pp. 212–219, 2016.

[7] B. Singh, P. Sihag, and K. Singh, "Modelling of impact of water quality on infiltration rate of soil by random forest regression," *Modeling Earth Systems and Environment*, vol. 3, pp. 999–1004, 2017.

[8] Y. Li, C. Zou, M. Berecibar, E. Nanini-Maury, J. C.-W. Chan, P. Van den Bossche, J. Van Mierlo, and N. Omar, "Random forest regression for online capacity estimation of lithium-ion batteries," *Applied energy*, vol. 232, pp. 197–210, 2018.

[9] F. Xiao, X. kang Wang, W. hui Hou, X. yang Zhang, and J. qiang Wang, "An attention-based bayesian sequence to sequence model for short-term solar power generation prediction within decomposition-ensemble strategy," *Journal of Cleaner Production*, vol. 416, p. 137827, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959652623019856

[10] I. Jamil, H. Lucheng, S. Iqbal, M. Aurangzaib, R. Jamil, H. Kotb, A. Alkuhayli, and K. M. AboRas, "Predictive evaluation of solar energy variables for a large-scale solar power plant based on triple deep learning forecast models," *Alexandria Engineering Journal*, vol. 76, pp. 51–73, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1110016823004908

[11] T. Cabello-López, M. Carranza-García, J. C. Riquelme, and J. García-Gutiérrez, "Forecasting solar energy production in spain: A comparison of univariate and multivariate models at the national level," *Applied Energy*, vol. 350, p. 121645, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261923010097

[12] I. Jebli, F.-Z. Belouadha, M. I. Kabbaj, and A. Tilioua, "Prediction of solar energy guided by pearson correlation using machine learning," *Energy*, vol. 224, p. 120109, 2021. [Online]. Available: https://www.sciencedirect.

com/science/article/pii/S0360544221003583

[13] N. Aksoy and I. Genc, “Predictive models development using gradient boosting based methods for solar power plants,” *Journal of Computational Science*, vol. 67, p. 101958, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877750323000182

[14] R. C. Deo, A. M. Ahmed, D. Casillas-Pérez, S. A. Pourmousavi, G. Segal, Y. Yu, and S. Salcedo-Sanz, “Cloud cover bias correction in numerical weather models for solar energy monitoring and forecasting systems with kernel ridge regression,” *Renewable Energy*, vol. 203, pp. 113–130, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S096014812201833X

[15] V. Chandran, C. K. Patil, A. Merline Manoharan, A. Ghosh, M. Sumithra, A. Karthick, R. Rahim, and K. Arun, “Wind power forecasting based on time series model using deep machine learning algorithms,” *Materials Today: Proceedings*, vol. 47, pp. 115–126, 2021, nCRABE. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214785321028388

[16] S. Pfenninger and I. Staffell, “The Renewables ninja,” http://wui.cmsaf.eu/safira/action/viewDoiDetails?acronym=SARAH_V001, 2016, accessed: 2023-10-01. [Online]. Available: http://wui.cmsaf.eu/safira/action/viewDoiDetails?acronym=SARAH_V001

[17] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, “Julia: A fresh approach to numerical computing,” *SIAM Review*, vol. 59, no. 1, pp. 65–98, 2017. [Online]. Available: https://epubs.siam.org/doi/10.1137/141000671

[18] M. Besançon, T. Papamarkou, D. Anthoff, A. Arslan, S. Byrne, D. Lin, and J. Pearson, “Distributions.jl: Definition and modeling of probability distributions in the juliastats ecosystem,” *Journal of Statistical Software*, vol. 98, no. 16, pp. 1–30, 2021. [Online]. Available: https://www.jstatsoft.org/v098/i16