Solar irradiation forecasting using Prophet

Todizara Andrianajaina[1], Tsivalalaina David Razafimahefa[1]

[1] Ecole Supérieure Polytechnique Antsiranana, University of Antsiranana, Antsiranana, Madagascar

Correspondent: todizara.andrianajaina@gmail.com

ABSTRACT

Solar irradiance prediction is critical for the scheduling of electricity output in photovoltaic (PV) power plants. To that goal, we developed algorithms for forecasting solar irradiation. We concentrate on the Prophet prediction algorithm. The research is based on data collected at the University of Antsiranana during a two-year period. The metrics of the approaches utilized have been validated, and the offered methods are effective in terms of predicting.

Keyword *solar irradiance, forecasting, prophet, machine learning*

1. INTRODUCTION

Large-scale electricity production is possible for a nation like Madagascar thanks to photovoltaic (PV) power plants. It is crucial for a photovoltaic power plant to forecast the level of solar irradiance for a specific time period in order to optimize operating costs through production scheduling. This is because the electricity generation from solar photovoltaic sources is directly proportional to solar irradiance. As a result, the goal of this research is to apply machine learning techniques to develop models that can accurately estimate solar irradiance while taking into consideration meteorological data [1-3].
A time series is made up of the sun radiation measurements made during a specific time period [4,5]. We used statistics to examine the seasonality, trend, and stationarity of the time series as well as the stationarity of the data we already had. We discovered that the data on solar irradiance shows seasonality across a one-year cycle. This shows that the position of the day in relation to the year affects the solar irradiation on a certain day of the year. Additionally, the sun irradiance is influenced by the weather, which may vary quickly.
We will learn more about Prophet forecasting in this post as well as how it may be used to forecast solar irradiance [6-8].

2. PROPHET FORECASTING

Working with time series data can be challenging and unpleasant, and the many techniques used to create models may be quite picky and challenging to fine-tune. If you are working with data that contains numerous seasonality, this is especially true. Traditional time series models, such as SARIMAX, also include a number of strict data requirements, including stationarity and uniformly spaced values. If you don't have a deep grasp of neural network design, other time series models like Recurring Neural Networks with Long-Short Term Memory (RNN-LSTM) may be quite complicated and challenging to work with. Time series analysis thus has a high entrance hurdle for the typical data analyst.

In order to address this, a few Facebook researchers published an article in 2017 titled "Forecasting at Scale" that launched the open-source project Facebook Prophet, which enables time-series modeling for data analysts and data scientists worldwide [9]. In fact, Prophet is a free, open-source toolkit (R and Python) for time-series data forecasting using an additive model. Even those without much experience in this sector may easily examine time series using this library.

The following issues are addressed by Facebook Prophet [10]:

- The challenge of developing trustworthy forecasting models: Because this field demands specialized training

- The inflexibility and fragility of computerized forecasting methods

The additive model is the base of Prophet. This indicates that a time series is modeled as the accumulation of several parts. This is referred to as a time series decomposition model.

The Additive Model is as follows:

$$y(t) = g(t) + s(t) + \varepsilon(t) \quad (1)$$

where

- y(t) is the time series model,
- g(t) the trend,
- s(t) the seasonal component,
- ε(t) the random or error component.

Prophet adds a new component that corresponds to the impact of vacations on the model.

Thus, the decomposition model of Prophet is the following:

$$y(t) = g(t) + h(t) + s(t) + \varepsilon(t) \quad 2)$$

where h(t) corresponds to the holiday effect (h as holidays).

2. RESULT AND DISCUSSION

A unit root test is a sort of statistical test that includes the Augmented Dickey-Fuller test. A unit root test is intended to assess the degree to which a time series is dominated by a trend. The Augmented Dickey-Fuller test may be one of the most popular unit root tests out of the many available. The information criteria is optimized across a range of various lag values using an autoregressive model [11].

The time series' unit root representation and lack of stationary behavior are the test's null hypotheses (has some time-dependent structure). The time series being steady is the alternative hypothesis (rejecting the null hypothesis).

If the null hypothesis (H0) cannot be ruled out, it is likely that the time series is non-stationary and has a unit root. Its structure is rather time-dependent. Alternate Hypothesis (H1): The null hypothesis is disproved, indicating that the time series is stationary since it lacks a unit root. It has no structure that changes with time. Using the test's p-value, we interpret the outcome. A p-value over the threshold indicates that we fail to reject the null hypothesis, whereas a p-value below a threshold (such as 5% or 1%) indicates that we do so (non-stationary).

p-value > 0.05: The data have a unit root and are non-stationary, failing to reject the null hypothesis (H0). If the p-value is less than 0.05, the data are stationary and do not have a unit root, rejecting the null hypothesis (H0).

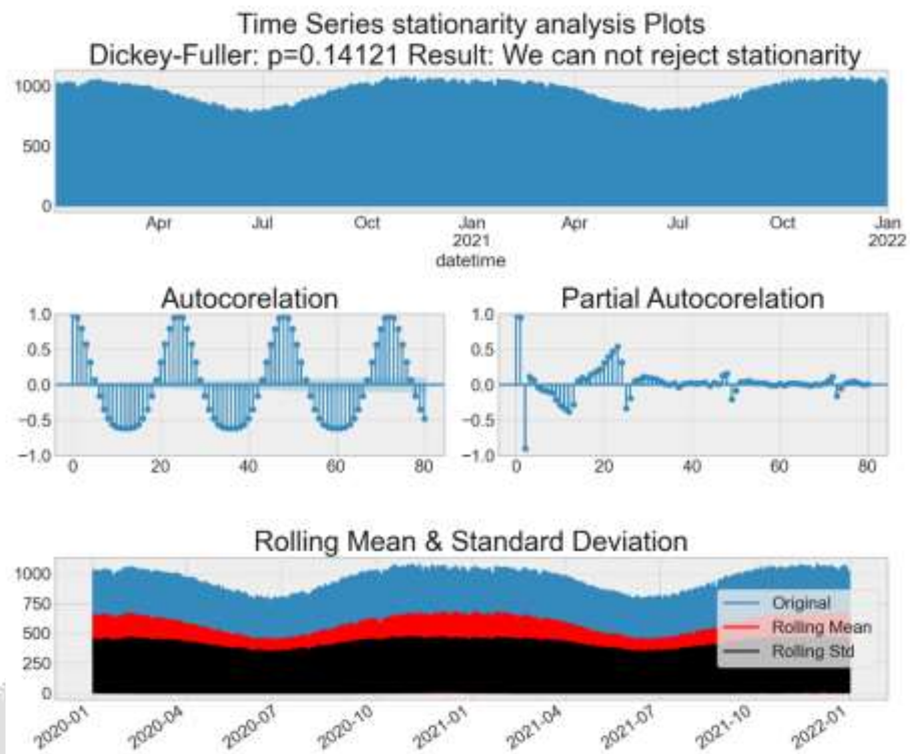

Fig -1 Irradiance Time series stationarity analysis

Irradiance time series have year seasonality. We will apply Prophet algorithm to forecast its value.

We have two years of data. We divided the data into three categories: learning, validation, and testing. We obtained the following result after applying the Prophet algorithm to estimate irradiance.

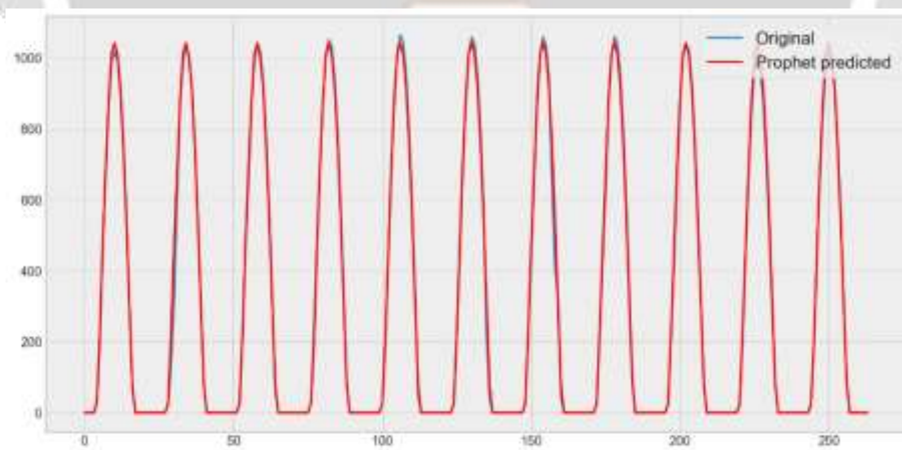

Fig -2 Irradiance predicted by Prophet

The algorithm performed well in predicting the Irradiance values in the future, as seen in the image. We will examine the Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R2 produced by the algorithm in order to better understand the importance of the forecast [12].

MAE refers to Mean Absolute Error, which is

$$MAE = \frac{1}{n}\sum_{1}^{n} |y_i - \hat{y}_i| \tag{3}$$

This gives less weight to outliers, which is not sensitive to outliers.

MAPE refers to Mean Absolute Percentage Error, which is

$$MAPE = \frac{100}{n}\sum_{i}^{n}\frac{y_i - \hat{y}_i}{y_i} \quad (4)$$

MAE = 90.358,
MAPE = 0.056,

Always with the aim of appreciating the quality of the prediction, we will display in a boxplot the residual (original - prediction).

Fig –3 Residual boxplot

We notice that the residuals are centered on zero (the average is equal to zero). The residuals are between -10 and 9kW/m^2.

3. CONCLUSION AND PERSPECTIVES

We presented a methodology for developing solar irradiance models that account for the yearly cycle's temporal dependency. Our research demonstrates that Prophet is capable of accurately forecasting solar irradiance. Even though the suggested approach is successful and can accurately predict the overall trend of solar irradiance, the prediction error can still be reduced by, for instance, accounting for exogenous values and by taking other seasonality, such as meteorological seasons, rather than just the years as we did.

Therefore, there is room for improvement. For instance, by training the model on data that contain observations on the same day of the year for all other years, the model may integrate long-term dependency (one year). Since the meteorological data we presently have only approximated the current meteorological conditions, we have utilized linear interpolation to fill in the gaps where observations are absent. If weather data is more precise, performance may be enhanced. Finally, we may take into account additional external data that may have a stronger link with solar irradiance, including solar activity or satellite photos of clouds.

6. REFERENCES

[1] Ahmed, R., Sreeram, V., Mishra, Y., & Arif, M. D. (2020). A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. *Renewable and Sustainable Energy Reviews*, *124*, 109792.

[2] Singla, P., Duhan, M., & Saroha, S. (2021). A comprehensive review and analysis of solar forecasting techniques. *Frontiers in Energy*, 1-37.

[3] Gupta, P., & Singh, R. (2021). PV power forecasting based on data-driven models: a review. *International Journal of Sustainable Engineering*, *14*(6), 1733-1755.

[4] Bawdekar, A. A., Prusty, B. R., & Bingi, K. (2022). Sensitivity Analysis of Stationarity Tests' Outcome to Time Series Facets and Test Parameters. *Mathematical Problems in Engineering*, *2022*.

[5] Suresh, S., Jain, A., & Suneeta, V. (2022). Time Series Analytics—A Fuel to Data Science in Forecasting Analytics for Time-Stamped Data. In *Rising Threats in Expert Applications and Solutions* (pp. 87-95). Springer, Singapore.

[6] Daraghmeh, M., Agarwal, A., Manzano, R., & Zaman, M. (2021, June). Time series forecasting using Facebook Prophet for cloud resource management. In *2021 IEEE International Conference on Communications Workshops (ICC Workshops)* (pp. 1-6). IEEE.

[7] Sivaramakrishnan, S., Fernandez, T. F., Babukarthik, R. G., & Premalatha, S. (2022). Forecasting Time Series Data Using ARIMA and Facebook Prophet Models. In *Big data management in Sensing* (pp. 47-59). River Publishers.

[8] Jha, B. K., & Pande, S. (2021, April). Time series forecasting model for supermarket sales using FB-prophet. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 547-554). IEEE.

[9] Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, *72*(1), 37-45.

[10] Asha, J., Rishidas, S., SanthoshKumar, S., & Reena, P. (2019, October). Analysis of temperature prediction using random forest and facebook prophet algorithms. In *International Conference on Innovative Data Communication Technologies and Application* (pp. 432-439). Springer, Cham.

[11] Harris, R. I. (1992). Testing for unit roots using the augmented Dickey-Fuller test: Some issues relating to the size, power and the lag structure of the test. *Economics letters*, *38*(4), 381-386.

[12] Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*, e623.