Case Study of Wind Energy Production Prediction using ARIMA Model

Hyuntae Park

Department of Chemical and Biological Engineering, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, Republic of Korea.

parksapphire@korea.ac.kr

ABSTRACT

Wind power is one of the most efficient and reliable renewable energy sources. For the effective use of wind power, the prediction of wind energy generation is an important task due to its stochastic nature. In this article, one day ahead forecast of Estonian wind energy production is made based on ARIMA model with seasonality. Comparison with the actual production indicated high errors in the forecast. Explanations for the deviation and potential solutions are presented. The data and jupyter notebook used can be accessed in the following link: github.

Keywords: Wind energy, Wind power forecasting, ARIMA

INTRODUCTION

Wind energy production fluctuates seasonally and throughout the day with chaotic turbulence. This makes the forecasting of wind power an important factor for designing a stable and reliable power system based on wind energy [1].

Statistical approach is a class of wind power production forecasting method. It is based on historical wind energy production data and involves the application of machine learning algorithms (i.e., neural network) such as ARIMA, Support Vector Machine (SVM) and Random Forest regression [2].

In this article, a case study of short-term (day ahead) forecast is performed based on statistical approach as an extension of Noman Shabbir et al. [3]. This article expands the scope of model selection by introducing ARIMA to forecast the next day's wind energy production. An evaluation is done for the accuracy of the selected method.

ARIMA MODEL

The ARIMA model is consisted of three parts: autoregressive (AR), moving average (MA), integrated (I). The AR part is used to describe the current value of a time-series, x_t, as a function of p past values. Since the AR model assumes linear dependence of the current value from the past values, x_t is expressed as a linear combination of order p [5].

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t \tag{1}$$

φ_i are the parameters of the AR model and ε_t denotes the white noise error term.

The MA model is an alternative to the AR model, linearly relating x_t to q past error terms.

$$X_t = \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{2}$$

For the MA model, θ_i are the parameters and ε_t, ε_{t-i} express the error term as in the AR part. Finally, the I part notates the degree of differencing. Since both the AR and MA models are applied to stationary data, differencing is implemented to delete the trend in the data.

$$Y_t = X_t - X_{t-1} \tag{3}$$

Y_t from equation 3 is obtained from first-order differencing of time series data X_t.

Putting three parts into one, the ARIMA model can be expressed as ARIMA(p, d, q). For data with seasonal trend, seasonal orders *P, D, Q* are additionally introduced with the seasonal period *m* and is expressed as SARIMA(p, d, q)(P, D, Q)[m] as the final form.

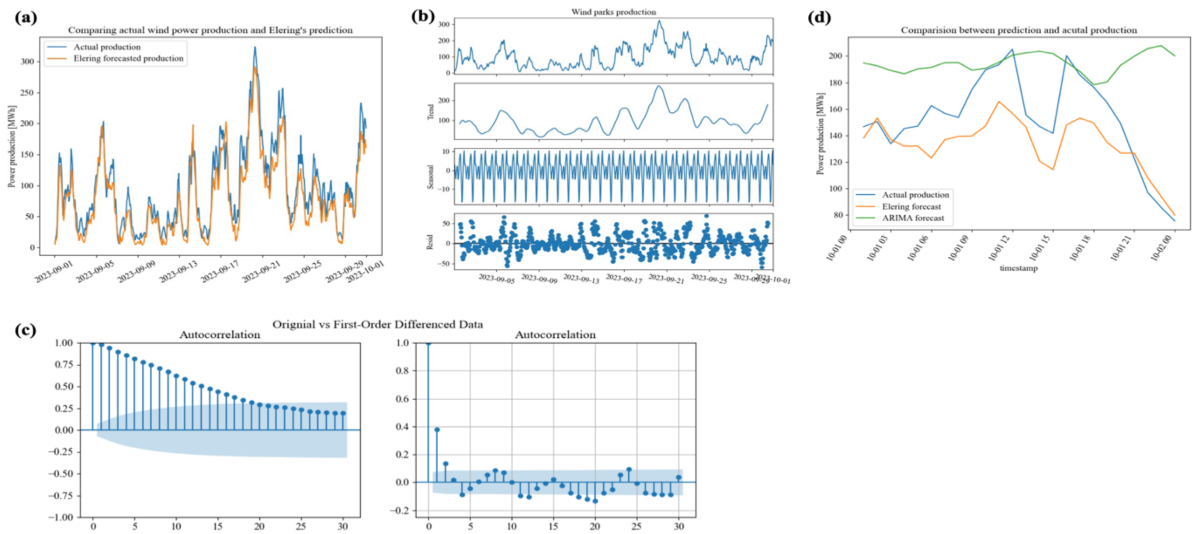
RESULTS AND DISCUSSION

The case study of wind power production forecasting was performed based on the data provided through Elering's website [4]. The wind power generated in Estonia for 30 days in September 2023 is shown in Figure 1(a) in comparison with the company's prediction.

Figure 1(b) shows the result of seasonal decomposition. As seen in the 'Seasonal' curve, the data clearly shows seasonality on a daily basis, making it reasonable to set the *m* to 24. Moreover, the autocorrelation plot of the original data in Figure 1(c) steadily decreases, implying non-stationarity. By applying first-order differencing, the autocorrelation plot exponentially decreases, indicating stationarity is reached. Thus, optimal *d* and *D* values are set to 1.

Using the python function auto_arima from the python library pmdarima.arima [6], the optimal orders of the non-determined parameters are fitted . The result of the search suggests SARIMA(2, 1, 2)(0, 1, 1)[24] as the optimal model with an aic score of 5534.63252.

Figure 1. Plots of the data analysis and forecasts



The SARIMAX function from the statsmodels.tsa.statespace.sarimax python library [7] is then applied to train optimal SARIMA model with the wind power production data. Predictions are made from the trained model for 24 hours ahead and is compared with the actual production and the forecast made by the corporation in Figure 1(d). The SARIMA model's prediction overestimates the power production and seems to weakly follow the tendency of the actual production until 10-01 18:00.

To improve the prediction, introducing a polynomial fit for the trend could be considered. The auto_arima search was based on a linear trend with time which does not perfectly match the trend shown in Figure 1(b). Also, addition of data preprocessing and model validation stages could enhance the accuracy. Finally, instead of making forecasts for 24 time stamps at once, a step-by-step prediction can decrease the error.

CONCLUSION

In this article, a statistical approach based on the ARIMA model is used to make forecasts of wind power production. An optimal model of SARIMA(2, 1, 2)(0, 1, 1)[24] was selected and trained. The prediction had a RMSE value of 55.6849 which is larger than that of Elering's algorithm: 26.1859 and Noman Shabbir et al. [3]'s prediction with SVM: 18.481.

The suggested reasons for this deviation are crude analysis of the trend and absence of data preprocessing and model validation steps. Potential improvements can be made by enhancing such factors and adapting a step-by-step prediction.

REFERENCE

[1] Demolli, H., Dokuz, A. S., Ecemis, A., & Gokcek, M. (2019). Wind power forecasting based on daily wind speed data using machine learning algorithms. *Energy Conversion and Management, 198*.

[2] Wang, X., Guo, P., & Huang, X. (2011). A Review of Wind Power Forecasting Models. *Energy Procedia*.

[3] N. Shabbir, R. AhmadiAhangar, L. Kütt, M. N. Iqbal and A. Rosin, "Forecasting Short Term Wind Energy Generation using Machine Learning," *2019 IEEE 60th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON)*, Riga, Latvia, 2019, pp. 1-4, doi: 10.1109/RTUCON48111.2019.8982365.

[4] Elering AS. *Wind parks*. Retrieved from Elering Live: https://dashboard.elering.ee/en/system/with-plan/production-renewable?interval=minute&period=days&start=2023-10-08T15:00:00.000Z&end=2023-10-09T14:59:59.999Z

[5] Shumway, R. H., & Stoffer, D. S. (2017). *Time Series Analysis and Its Applications: With R Examples.* Springer.

[6] Smith, Taylor G., *et al*. pmdarima: ARIMA estimators for Python, 2017-, http://www.alkaline-ml.com/pmdarima [Online; accessed 2023-10-25]

[7] Seabold, Skipper, and Josef Perktold. "statsmodels: Econometric and statistical modeling with python." *Proceedings of the 9th Python in Science Conference.* 2010.