This is a peer-reviewed, author's accepted manuscript of the following conference proceedings output:
Sathler, K. P. B., & Kolios, A. (2022). The use of machine learning and performance concept to monitor and predict wind power output. In 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET) (pp. 1-8). IEEE. https://doi.org/10.1109/icecet55527.2022.9873076

The Use of Machine Learning and Performance Concept to Monitor and Predict Wind Power Output

Kelvin Palhares Bastos Sathler
NAOME Department
University of Strathclyde
Glasgow, Scotland
kelvin.palhares-bastos-sathler@strath.ac.uk

Athanasios Kolios
NAOME Department
University of Strathclyde
Glasgow, Scotland
athanasios.kolios@strath.ac.uk

***Abstract*—Monitoring and predicting wind power output more precisely can be very beneficial for an increasingly competitive Wind Power industry. Although many advances have been made throughout the last decades, the production forecast is still based mainly on the manufacturing power curve and wind speed. Even though this approach is very useful, especially during the design phase, it does not consider other factors that affect production, such as topography, weather conditions, and wind features. A more precise prediction model that is able to recognize production fluctuation and is tailored using current operational data is proposed in this paper. The model analyzes the performance through Meteorological Mast Data (Met Mast Data) and then uses it as an input to monitor and predict power output. As a result, the model proposed achieves high accuracy and can be key to understanding the wind turbine asset's behavior throughout its lifespan, assisting operators in decision making to increase overall power production.**

***Index Terms*—wind power curve, output prediction, performance, met mast data, machine learning, monitoring**

I. Introduction

Wind Power has consistently increased its importance in the energy industry, achieving 15% of Europe's electricity demand in 2021 [1]. A more competitive market has enhanced the need for a better performance with higher production at lower costs. With that in mind, improving wind power output prediction has become important not only in its design, but also in its operational phase. A better understanding of production fluctuations could be beneficial to operators in order to minimize curtailment losses [2], plan maintenance in better periods, considering lower wind regimes [3], or even assist them to identify abnormalities in the equipment [4], [5]. All this together, could support operators to increase the overall production as well as evaluate more precisely the real efficiency of the wind farm.

The most common way to monitor and predict wind power output is through the power curve, which in most cases is provided by the manufacturer. As defined by BS IEC 61400-12-1-2017, the power curve is based on the average power produced in a predetermined wind speed bin [6]. Although very useful, these curves do not usually consider the external features and some of the possible operational losses. Therefore, during the design phase a rate is considered in order to calculate the net production. According to Ioannou et al. [7] 90% is a reasonable estimation to be used in the design phase. However, in the operational phase, this approximation does not help operators to understand and identify what is causing the operational losses and fluctuations in production.

Wind energy output is mainly calculated through the amount of Kinect energy flux from the wind taken by the rotor, considering the density of the air, wind speed, rotor area and the power coefficient. The power coefficient is what determines how much energy can actually be captured from the wind and it is related to some wind features and rotor features, which include the tip speed ratio and blade pitch angle [8]. According to Betz's law, the theoretical maximum possible performance is equal to 16/27, i.e. 59.3% of the kinetic energy in wind.

External factors such as climate, wind conditions and topography can clearly affect the outcome and could be the reason for high fluctuation. Recent studies have been trying to create alternative curves to increase accuracy in prediction and comprehension of production. In the literature there are studies creating curves adding more inputs, such as air density [9], humidity [10], wind direction [11], turbulence [12], and periods of the day [13]. Also machine learning has been largely used to predict wind power output, as shown in [14]–[20] and also to create a model of day-ahead prediction [21]. Even though these models are very useful and beneficial for operators, the great number of curves can make decision making more complex, and also, they do not seem to offer a totally tailored approach.

A solution given by Sathler et al. [22] is the creation of a new variable called performance, where the variance of production is calculated by dividing each value by the maximum registered on the related bin. As a result, the index itself has proved to have similar efficiency in predicting production and, when used together with wind speed, the results were higher than 96% when considering the entire farm. The results also showed that some periods with similar wind can have drastically different production when relying on the index created. Even though these are promising results, the study does not bring solutions for monitoring performance in an effective way.

As mentioned before some of possible reasons for fluctuation in production are due to external factors. To monitor the external factors, farms have a separate tower with sensors in order to measure several wind and climate features. Therefore, the goal of this study is to check if the information provided by this tower, also known as Meteorological Mast Data (Met Mast Data), can be used to predict the performance index of the turbines and check if they provide a reliable power output prediction to assist operators in monitoring their turbines.

The rest of this paper is organized as follows. Chapter 2 describes briefly the methods used in this study, while in Chapter 3, a description of the data and simulation strategies are provided. Chapter 4 contains the results and discussion, followed by Chapter 5 which contains the conclusion and suggestions for future work.

II. Methods

Machine learning has become very popular in recent years; this is due mainly to the advances in technology, especially those related to storage systems and faster processing of data. There are many types of algorithms and methods inside machine learning, which can help to find patterns in their outcomes, connecting inputs to outputs to make predictions with lower errors. In this study, regression-supervised models will be used. They are classified as supervised because the algorithm is trained based on historical data previously available, and as regression since production and performance are considered quantitative continuous values.

To design both models, production output and performance prediction, four different methods were chosen: Linear Regression (LNR), Decision Tree Regression (DTR), Support Vector Regression (SVR) and Random Forest Regression (RFR). Table I gives a short explanation on the methods with references for further interest. The goal of this paper is to keep the models as simple as possible and check if the model is consistent, avoiding the need of high computational resources. For this reason more advanced methods were not assessed.

A normal problem in machine learning is the selection of the data, which can affect results positively or negatively. So, to avoid any misleading in the conclusions or any bias due to data selection, k-fold Cross Validation (CV) will be used. In this method, the data are split into "k" equal parts, where each of these parts is used as a test set, while the rest of data is used as a training set, so the model runs k-times. This process provides "k" different outcomes, so further statistical analysis will lead to evaluating each regression model, and the one that fits best for the purpose of this study will be selected.

Finally, to assess the accuracy of the method proposed, four different metrics were selected. First, the coefficient of determination, or R-squared (R^2), measures how well a dependent variable explains the independent variable. This metric is calculated considering the squared sums of the "distance" between predicted values and the actual observed ones and the total sum of squares, according to Eq. 1.

TABLE I: Description of Machine Learning Methods Selected.

Model	Description	Reference
Linear Regression (LNR)	Linear prediction method. Look for best fit, with lower errors, considering straight linear equation.	-
Decision Tree Regression (DTR)	It is a nonlinear supervised prediction method, which creates conditional statements.	[23]
Support Vector Regression (SVR)	Based on Support Vector Machines, uses hyperparameters with a tolerance to minimize errors.	[24]
Random Forest Regression (RFR)	Construction of multiple decision trees. A probabilistic analysis is made among those decision tree to select the best prediction.	[25]

$$\mathrm{R}^2 = 1 - \frac{\mathrm{SSR}}{\mathrm{SST}} \tag{1}$$

where SSR is the sum of squared residuals and SST total sum of squares. The results vary from 0 to 1, where 1 means perfect correlation, in other words, the dependent variables explain 100% of the variance in the independent variable. Conversely, a result equal to zero means there is no correlation.

The other three metrics calculate the mean value of the residuals, they are: the Mean Absolute Error (MAE), the Mean Percentage Absolute Error, and the Root Mean Squared Error (RMSE) As suggested by its name, MAE (Eq. 2) considers the mean absolute value of the difference between the prediction and the real value. Similar to MAE, MAPE (Eq. 3) also consider absolute values, however, the error is divided by the real value, providing a ratio of its accuracy. Finally, by being squared, RMSE (Eq. 4) penalizes bad predictions, providing important information from the accuracy of the model. In Eqs. 2 to 4, y_i is the tested value and $\hat{y}_i$ is the predicted one.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{2}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{3}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{4}$$

III. Data Analysis

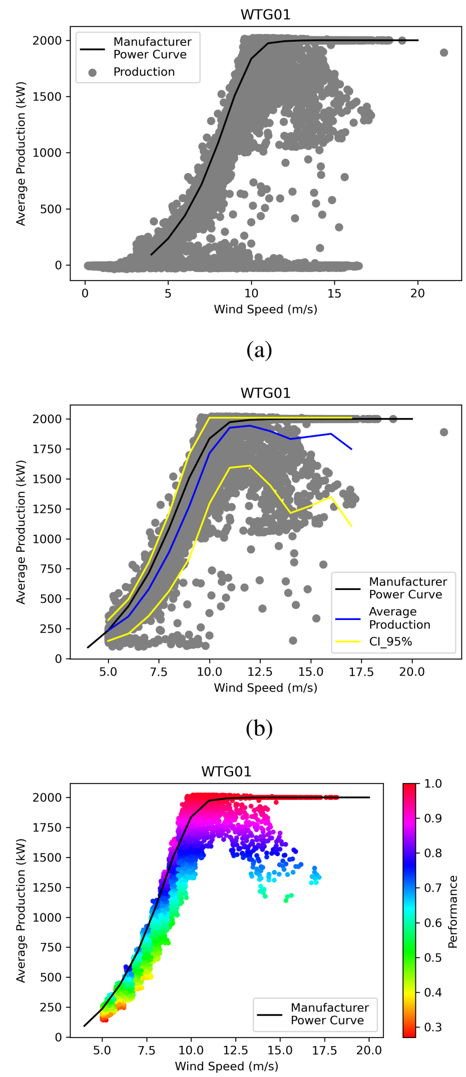
For confidentiality reasons the location and model cannot be disclosed. To test the model proposed, SCADA data from five wind turbines of 2MW and Met Mast Data from the farm tower were used. It was provided information from 172 days with 10 minutes range, totaling 24,768 records per turbine. From the SCADA data was retrieved production and wind speed. The Met Mast Data, on the other hand, includes wind speed measured at five different heights, direction of the wind at four different points, humidity and temperature, both from two distinct points, and finally ambient pressure.

A. Pre-processing

From these data some abnormalities, such as negative outputs, missing data, and periods where wind speed was out of the production range were removed. Even though the cut-in speed of this turbine is 4 m/s, recordings below 5 m/s were not considered, because of the high fluctuations caused by the starting-up of the turbine. Production below 100kW was eliminated as well, since this is the minimum output according to the manufacturer curve. Hence, this values are more likely to be an error and could mislead the model. The goal of this work is to create a model to predict production and performance. Therefore, outliers can minimize the accuracy of the model. Considering that, an interval of confidence of 99% was calculated and the values out of this range were considered to be outliers.

As mentioned in the introduction, the performance index was calculated by the division of each validated data to the maximum production recorded in the same bin. The wind speed range of 0.5m/s was selected for this study, following the recommendation from BS IEC 61400-12-1-2017. From now on, all the processes will be presented only considering the first wind turbine generator (WTG1) in order not to be repetitive. But the process explained here is the same applied to all turbines and in the same farm. As a result, from the original 24,768 recordings provided, 15,141 were used to assess the methodology proposed after pre-processing in WTG1.

Figure 1 illustrates the pre-processing evolution in three different stages. Figure 1a represents the total amount of data provided; it can be seen that the production fluctuates significantly during the period assessed. In Figure 1b, the first criteria for data reduction, wind speed below 5 m/s and production below 100kW, was undertaken and the interval of confidence of 99% was calculated. In this graph it is clear how outliers, especially those above the upper limit, can affect results. Since the performance is a rate between each value and its maximum, the outliers would affect all values in that bin, creating distortions in the model. It is important to note that this outlier was occasional, since there are no other points around it in the range, which justifies its removal. Finally, the Figure 1c includes a colour map with the performance index calculated. The black line is the power curve provided by the manufacturer and the blue one is the average production.



(a)

(b)

(c)

Fig. 1: Pre-processing steps in WTG1: a) Complete Data; b) First removal criteria + Interval of Confidence; c) Data Pre-processed + Performance.

B. Model Selection

In the model selection activity, basic procedures and parameters were used. Nonetheless, a deep investigation of parameters was developed to run the methodology proposed and it is presented in Sub-Section III-C. Hence, in few words, to find the best LNR model, Ordinary Least Square method was used. To train the RFR, one hundred trees was used and the kernel selected to the SVR was the “Radial Basis Function”. In all methods square error was used as a metric to the loss function and the input features were scaled through Standardization.

To avoid bias, a cross validation was performed to identify and select the best regression method. In this analysis it was considered k = 10 and the metric used was R^2. This study is divided into two different models: one to calculate the power output considering wind speed and performance (MODEL 1), and the other to predict the performance of production using Met Mast Data (MODEL 2). However, before defining the models, to evaluate the advantages of adding performance as an input, the same procedure was done considering only wind speed, which is illustrated in Figure 2. The graph presents a box plot of the result. The mean value and the standard deviation from each model is presented within the plot.

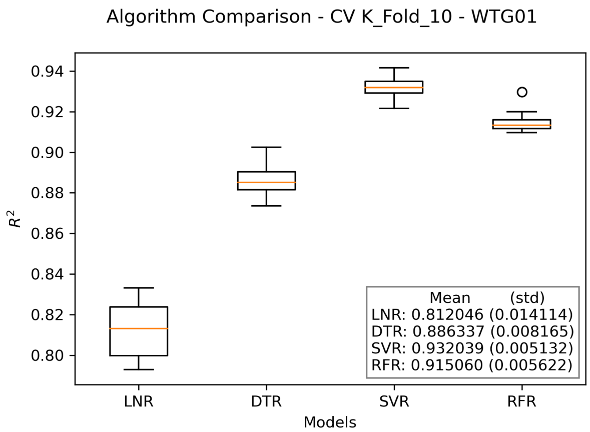


Fig. 2: Box Plot of CV of Power Output Prediction with only Wind Speed.

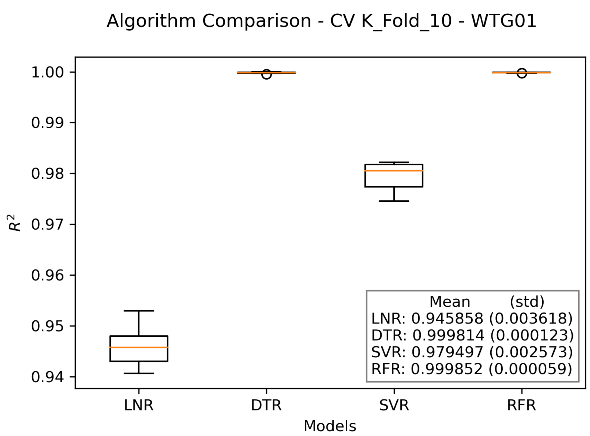


Fig. 3: Box Plot of CV of Power Output Prediction with Performance and Wind Speed (MODEL 1).

Figure 3 presents the results of the models to calculate production considering wind speed and performance. Even though considering only wind speed has some good values, the addition of performance significantly improved the accuracy in all methods. The high result is not a surprise since the performance index identifies the fluctuation of production in a certain bin, in other words, where exactly the production output will be. Although the visible curvature, LNR obtained a high result, around 0.94, which illustrates and reaffirms the importance of the performance index as an input. DTR and RFR, considered nonlinear regression conditional models, achieved nearly perfect correlation, with RFR having a slight better result.

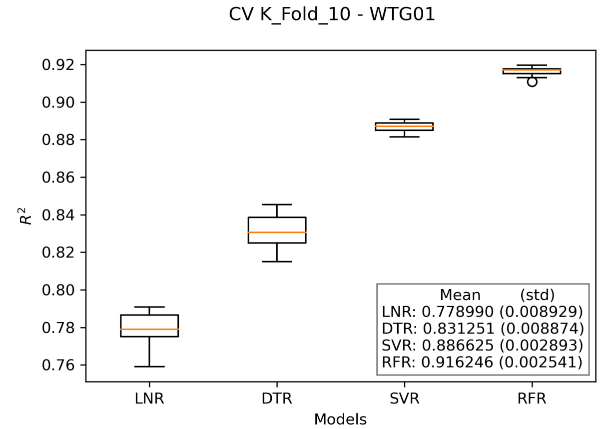


Fig. 4: Box Plot of CV Performance Prediction through Met Mast Data Prediction Model (MODEL 2).

The biggest challenge of the model proposed is the prediction of the performance index. Figure 4 shows the results of MODEL 2, where the performance was predicted using Met Mast Data as input. In this scenario SVR and RFR had the best results, achieving an average R^2 of 0.886 and 0.916 respectively. Considering the best outcome, RFR, the result is very consistent since the standard deviation was lower, around 0.0025, which means that around 91% of wind turbine performance can be explained by the Met Mast Data or external interference.

C. Power Output Model

As shown in the model selection analysis, RFR had performed better in both models proposed, so this method will be used to predict the performance index (MODEL 2) and to calculate the power output (MODEL 1). The possibilities behind RFR architecture are infinite. Besides the amount of ‘trees’ that can be analysed in each ‘forest’, parameters such as number of leafs, depth, and nodes of the architecture of the ‘trees‘ can also affect the results. More information about the parameters can be found in [27]. The strategy used in this paper to tune the best parameters for the models was divided into two parts. First, from a larger possibility of parameters, a number of random selected combination was calculated. Then, according to the best results in the random selection, a more strict range was retested, but this time considering all possibilities. A five fold cross validation was considered in each iteration. Table II shows the parameters and criteria assessed in each stage, as well as the one selected. Figure 5 shows the box plot of MODEL 2 after the parameters were redefined, improving the average R^2 by 0.54%. Since MODEL 1 achieved R^2 equals to 0.999, the basic structure was kept without further analysis.

Finally, from the total data provided, 30% was separated to validate the model, while 70% was used to train the

TABLE II: RFR Optimization Routine - WTG01.

Param.	Random	Single	Selected
Bootstrap	['True', 'False']	['False']	['False']
Max. Depth	None and linspace(20, 110, 4)	['None', 20, 80]	['None']
Max. Feature per leaf	[1, 2, 5, 10, 14]	[1, 2, 5]	[1]
Min. Samples per leaf	[1, 2, 4]	[1, 2]	[1]
Min. Samples Split	[2, 4, 10]	[2, 10]	[2]
Number of Estimators	linspace(10, 1010, 11)	[110, 210, 410]	[210]
Total Possible Combinations	4950	108	-
Total Tested	1000	108	-

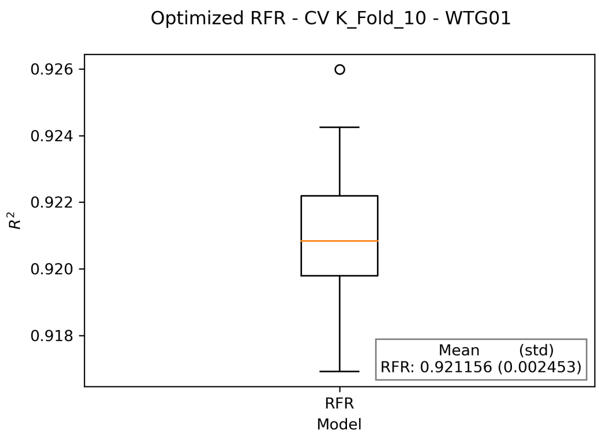


Fig. 5: Box Plot of MODEL 2 after RFR Optimization.

models. It is important to mention that the validation data were separated before the performance calculation. The goal behind this strategy is to emulate a real scenario, where the performance index is calculated only with provided data and the performance index predicted is unknown. It is expected this will increase the reliability on the models proposed and avoid any possible bias. In addition, to increment the results, the algorithm created was run five times with a random selection of the data set. Figure 6 summarizes the process proposed in this work. It is important to mention that to calculate the complete farm, the whole process was redone adding a new input variable to identify each one of the five turbines. In other words, “All Farm” scenario is not the average analysis of the five turbines individually, but it is a new simulation.

IV. Results and Discussion

The proposed approach proved to be efficient to predict the wind power output. Table III presents the results from the five iterations considering turbine WTG01, where the predictions achieved an average of 0.24% error when compared to the real production. Regarding the metrics, the results were also very consistent, independently of the iterations. The R^2 was close to 1.0 and the average RMSE was around 77.9kW; since this is a 2MW turbine, this error is acceptable considering the benefit the model can bring. Gross production, i.e., the one considering the manufacturer power curve, was presented as a reference as well as the net production, considering 90.0% performance.

It is important to note that the net production has achieved a good result as well; however, this value is more appropriate for the design phase. By using a fixed rate, this value does not help the operators to understand the production behavior, during the operational phase. As mentioned before, there are many factors that affect the turbine performance, many of which are related directly to the external factors and climate features. Therefore, the use of Met Mast Data to monitor performance and estimate production can be considered reliable. Some of the known operational losses are due to turbulence, air density, or wake effects, for example, and they can be linked to the differences in wind speed in different points, temperature, and wind direction, respectively, and all this information is provided by the Met Mast Data.

Another advantage of the model is that it gives extra information to operators about possible losses in the equipment’s efficiency. Components of the turbine tend to reduce its performance before breakdowns [26], but considering the high fluctuation in production, it is hard to identify if and when this possible loss is due to wind features or an equipment issue. Monitoring the expected performance through Met Mast Data could work as a reference to operators to check if the fluctuation is normal, considering the environment characteristics, or if this could be a mechanical or electrical problem.

Although it is possible to find in the literature better results in terms of accuracy, the model presented in this paper has the advantage of using less computational resources when compared to more advanced techniques, such as Artificial Neural Network. Also, since the model here proposed uses data from different sources, it can avoid redundancies and provide a new independent input for operators to monitor the real performance of the turbine. Furthermore, if more data are taken into account considering a whole year when training the model and an in depth evaluation for elimination of abnormalities is performed, an improvement in the performance prediction is expected. In this study, fault alarm data were not provided by the operators, which prevented a proper abnormalities evaluation.

While the model seems very useful to monitor production and performance in real time, future predictions were not evaluated until here. Knowing that it is very unlikely to have accurate forecasts within a 10 minute range, as calculated so far, an additional simulation of performance prediction considering average daily results was done and the results are shown in Figure 7. From the 172 days of data provided, the

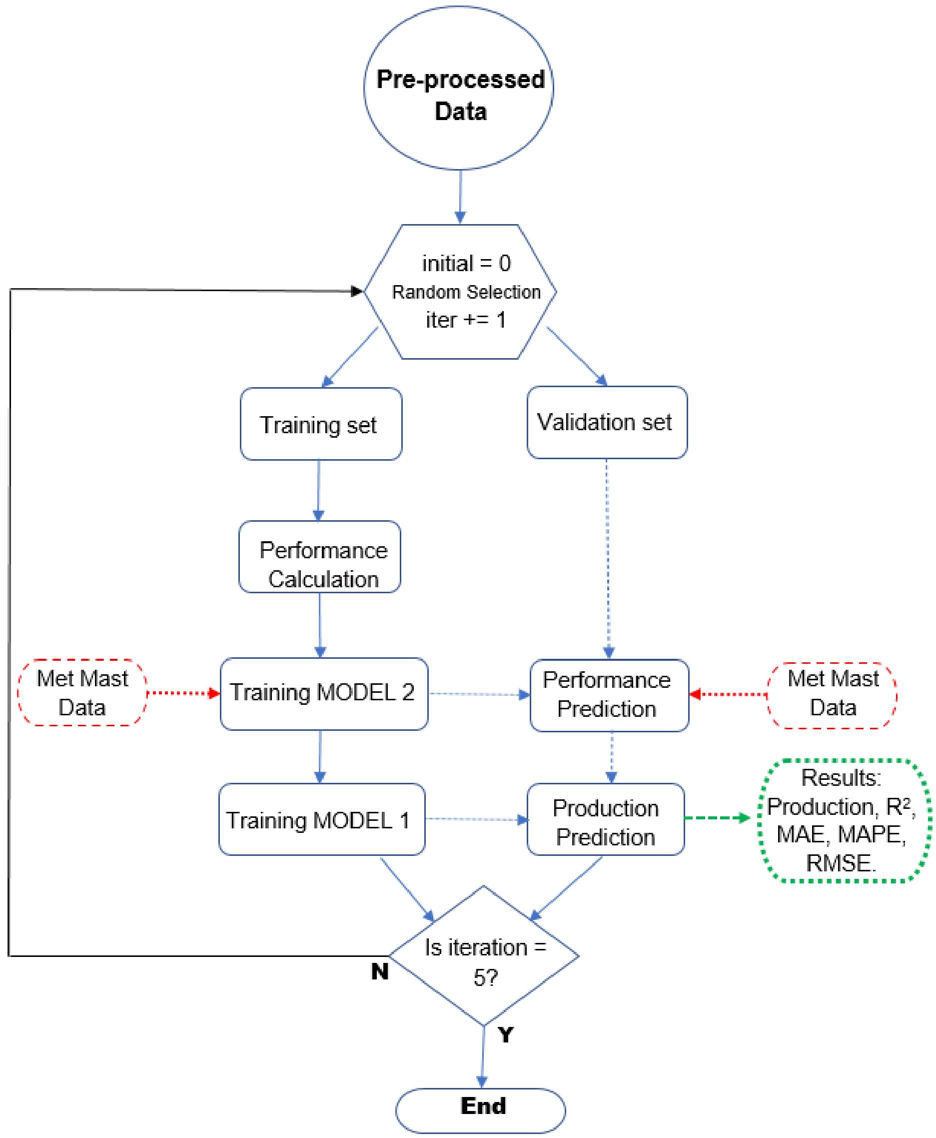


Fig. 6: Flowchart of the model proposed.

TABLE III: Results from WTG1 with 15,141 data points (RP = Real Production, MP = Model Prediction, GP = Gross Production, NP = Net Production)

Iteration		RP (kWh)	MP (kWh)	GP (kWh)	NP (kWh)	R^2	MAE	MAPE	RMSE
1	Values	883,364.64	885,941.83	972,915.38	875,623.84	0.9852	52.6931	6.00%	78.1422
	Δ	-	0.29%	10.14%	-0.88%				
2	Values	884,423.00	886,833.65	976,478.52	878,830.67	0.9848	53.4556	5.97%	78.6284
	Δ	-	0.27%	10.41%	-0.63%				
3	Values	888,095.65	890,885.79	978,283.22	880,454.90	0.9851	52.9711	5.94%	78.3595
	Δ	-	0.31%	10.16%	-0.86%				
4	Values	881,308.79	882,689.29	971,379.26	874,241.33	0.9854	52.2486	5.85%	77.2820
	Δ	-	0.16%	10.22%	-0.80%				
5	Values	889,771.90	891,025.18	977,694.58	879,925.12	0.9857	51.4527	5.77%	77.1723
	Δ	-	0.14%	9.88%	-1.11%				

month of January, i.e., 31 days, was separated out as a test set, and the other 141 days as training. In this simulation the R^2 of the performance prediction was 0.89 and RMSE was 0.04, which means the model can provide a good accuracy even considering larger ranges. This can be useful, especially in order to plan maintenance in advance, since in some periods the total production can be lower, even though the wind speed is the same.

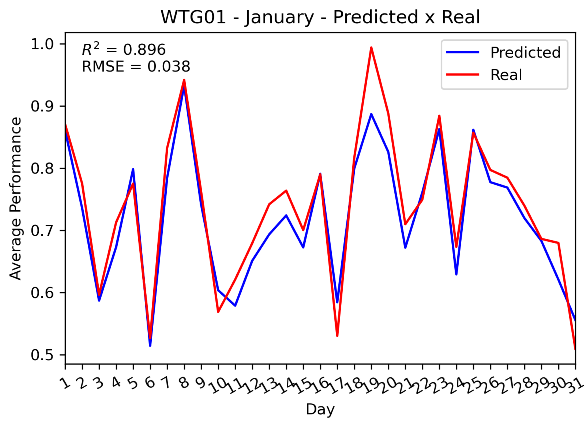


Fig. 7: Performance Prediction x Real Performance – WTG1 – January.

To sum up, Table IV presents the average results achieved in each turbine and when the entire farm is assessed together. Considering all simulations the average error between the real production and from the model was 0.16%. The MAPE and the RMSE was respectively, 5.65% and 72.42kW. The entire farm obtained a slight better results whrn compared to the individual analysis, with an average RMSE of 61.6kW. This could be due to the amount of data or by the identification of the turbine as an extra input. It is expected that the difference in individual performance can be better tracked when they are assessed together, especially considering the wake effect loss.

TABLE IV: Average results from all simulations.

Param.	WTG1	WTG2	WTG3	WTG4	WTG5	All
Error	0.24%	0.07%	0.16%	0.08%	0.18%	0.21%
R^2	0.9852	0.9871	0.9863	0.9860	0.9858	0.9905
MAE	52.5642	49.4654	49.5505	51.1040	50.8078	41.4567
MAPE	5.90%	5.54%	5.71%	5.83%	5.93%	5.00%
RMSE	77.9169	71.1578	73.7286	74.6535	75.4473	61.5918

V. Conclusion

The competitiveness of the energy market leads operators to make the best of its equipment with the lowest possible cost and a deep understanding of the production behavior can be a very strategic ally. The model proposed focused on relates Met Mast Data to the performance of production, or fluctuations, and uses this information as a new input to predict wind power output. The model proved to have a better accuracy when compared to the more traditional use of a manufacturer power curve. Considering all outcomes from all turbines, the R^2 results were superior to 0.98 in all individual iterations and obtained an average error of 0.14% when compared to the real production, the average R2, MAE, MAPE, and RMSE was respectively, 0.986, 50.7kW, 5.78%, and 74.6kW. Considering the ”all farm” simulation, the error, R2, MAE, MAPE and RMSE were 0.21%, 0.990, 41.5kW, 5.00% and 61.6kW, respectively.

The use of Met Mast Data to predict performance can be very beneficial: Firstly, because they contain most of the external information that directly affects wind power production, so it could reduce the use of several curves, which assess these inputs individually, to only one model. Secondly, for the authors, the turbine behavior is unique, which means that although known, the influence of each factor on production losses can vary from farm to farm due to the topography, wind conditions and climate. So, the use of the Met Mast Data related to real performance provides an increasingly tailored model. In addition, a daily average model was created and proved to be as effective as using a 10 minute range, which means the model can be used for short- or medium-term predictions, depending on the accuracy of climate and wind feature forecasts or a historical database.

To end up, the model proved to be very effective in helping operators with decision making as a tool to monitor performance and predict production in real time or for future predictions. As future work is recommended a more detailed feature analysis and reduction, eliminating possible disturbances. Also, the use of similar methodology with more advanced machine learning techniques could improve performance analysis. Finally, another important suggestion is to relate and quantify the losses in the performance with each input, through a “feature importance analysis”.

Acknowledgment

The authors thank the financial and technical support from ational Counsel of Technological and Scientific Development (CNPq) and Petróleo Brasileiro S.A. (Petrobras).

References

[1] WindEurope.: Wind energy in Europe: 2021 Statistics and the outlook for 2022-2026. (2022) http://www.windeurope.org

[2] R. Davison-Kernan, X. Liu, S. McLoone, and B. Fox.: Quantification of wind curtailment on a medium-sized power system and mitigation using municipal water pumping load, Renew. Sustain. Energy Rev., vol. 112, no. June 2019, pp. 499–507 (2019). 10.1016/j.rser.2019.06.004

[3] N. Y. Yürüşen, P. N. Rowley, S. J. Watson, and J. J. Melero.: Automated wind turbine maintenance scheduling, Reliab. Eng. Syst. Saf., vol. 200 (2020). 10.1016/j.ress.2020.106965

[4] Q. Huang, Y. Cui, L. B. Tjernberg, and P. Bangalore.: Wind Turbine Health Assessment Framework Based on Power Analysis Using Machine Learning Method, Proc. 2019 IEEE PES Innov. Smart Grid Technol. Eur. ISGT-Europe 2019 (2019). 10.1109/ISGTEurope.2019.8905495

[5] B. Karadayi, Y. Kuvvetli, and S. Ural.: Fault-related Alarm Detection of a Wind Turbine SCADA System, 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) (2021). 10.1109/HORA52670.2021.9461331

[6] L. T. Paiva, C. Veiga Rodrigues, and J. M. L. M. Palma.: Determining wind turbine power curves based on operating conditions, Wind Energy, vol. 17, no. 10, pp. 1563–1575 (2014). 10.1002/we.1651

[7] A. Ioannou, A. Angus, and F. Brennan.: Parametric CAPEX, OPEX, and LCOE expressions for offshore wind farms based on global deployment parameters, Energy Sources, Part B Econ. Plan. Policy, vol. 13, no. 5, pp. 281–290 (2018). 10.1080/15567249.2018.1461150

[8] S. Heier.: Grid Integration of Wind Energy (2014).

[9] R. K. Pandit, D. Infield, and J. Carroll.: Incorporating air density into a Gaussian process wind turbine power curve model for improving fitting accuracy, Wind Energy, vol. 22, no. 2, pp. 302–315 (2019). 10.1002/we.2285

[10] S. H. Danook, K. J. Jassim, and A. M. Hussein.: The impact of humidity on performance of wind turbine, Case Stud. Therm. Eng., vol. 14, no. April, p. 100456 (2019) 10.1016/j.csite.2019.100456

[11] C. Yan, Y. Pan, and C. L. Archer.: A general method to estimate wind farm power using artificial neural networks, Wind Energy, vol. 22, no. 11, pp. 1421–1432 (2019). 10.1002/we.2379

[12] Y. M. Saint-Drenan et al.: A parametric model for wind turbine power curves incorporating environmental conditions, Renew. Energy, vol. 157, pp. 754–768 (2020). 10.1016/j.renene.2020.04.123.

[13] L. Tian, Y. Song, N. Zhao, W. Shen, T. Wang, and C. Zhu.: Numerical investigations into the idealized diurnal cycle of atmospheric boundary layer and its impact on wind turbine's power performance, Renew. Energy, vol. 145, pp. 419–427 (2020). 10.1016/j.renene.2019.05.038

[14] A. Kusiak, H. Zheng, and Z. Song.: On-line monitoring of power curves, Renew. Energy, vol. 34, no. 6, pp. 1487–1493 (2009). 10.1016/j.renene.2008.10.022

[15] R. K. Pandit, D. Infield, and A. Kolios.: Gaussian process power curve models incorporating wind turbine operational variables, Energy Reports, vol. 6, pp. 1658–1669 (2020). 10.1016/j.egyr.2020.06.018

[16] U. Datta, J. Shi, and A. Kalam.: Primary frequency control of a microgrid with integrated dynamic sectional droop and fuzzy based pitch angle control, Int. J. Electr. Power Energy Syst., vol. 111, no. April, pp. 248–259 (2019). 10.1016/j.ijepes.2019.04.001

[17] P. P. Reboucas Filho, N. De Medeiros Mendonca E Nascimento, S. S. Araujo Alves, S. Luz Gomes, and C. Marques De Sa Medeiros.: Estimation of the energy production in a wind farm using regression methods and wind speed forecast, Proc. - 2018 Brazilian Conf. Intell. Syst. BRACIS 2018, pp. 79–84 (2018). 10.1109/BRACIS.2018.00022

[18] Y. Zhang, and Y. Zhao.: Research on Wind Power Prediction Based on Time Series, IEEE International Conference on Artificial Intelligence and Industrial Design (AIID), pp. 78–82 (2021). 10.1109/AIID51893.2021.9456531

[19] M. Mansoor, Q. Ling, and M. H. Zafar.: Short Term Wind Power Prediction using Feedforward Neural Network (FNN) trained by a Novel Sine-Cosine fused Chimp Optimization Algorithm (SChoA), 5th International Conference on Energy Conservation and Efficiency (ICECE), pp. 1-6 (2022). 10.1109/ICECE54634.2022.9758965

[20] S. Diop, P. S. Traore, and M. L. Ndiaye.: Wind Power Forecasting Using Machine Learning Algorithms, 9th International Renewable and Sustainable Energy Conference (IRSEC), pp. 1-6 (2021). 10.1109/IRSEC53969.2021.9741109

[21] L. Duchesne, E. Karangelos, A. Sutera, and L. Wehenkel.: Machine learning for ranking day-ahead decisions in the context of short-term operation planning, Electr. Power Syst. Res., vol. 189, no. April, p. 106548 (2020). 10.1016/j.epsr.2020.106548

[22] K. P. B. Sathler, A. Kolios, S. Al-Sanad, and J. Parol.: Application of the Overall Equipment Effectiveness Concept in Wind Energy Assets (2020). 10.3850/978-981-14-8593-0

[23] J. R. Quinlan.: Simplifying Decision Trees, Int. J. Hum. Comput. Stud., vol. 27, pp. 221–234 (1987). 10.1006/ijhc.1987.0321

[24] H. Drucker, C. J. C. Surges, L. Kaufman, A. Smola, and V. Vapnik.: Support Vector Regression Machines, Adv. Neural Inf. Process. Syst., vol. 1, pp. 155–161 (1997).

[25] L. Breiman.: Random forests, Mach. Learn., vol. 45, pp. 5–32 (2001). 10.1201/9780429469275-8

[26] M. Reder, N. Y. Yürüşen, and J. J. Melero.: Data-driven learning framework for associating weather conditions and wind turbine failures, Reliab. Eng. Syst. Saf., vol. 169, no. January 2017, pp. 554–569 (2018). 10.1016/j.ress.2017.10.004

[27] S. Saxena.: A Beginner's Guide to Random Forest Hyperparameter Tuning. (2020) https://www.analyticsvidhya.com/blog/2020/03/beginners-guide-random-forest-hyperparameter-tuning/