31761 - Renewables in Electricity Markets

Renewable energy forecasting
Let's compete!

Eléa Foulgoc s192478

Model construction and programming /Report writing: Digging into data, Results analysis, Lead Time resolution.

Théo Dullin s192446

Model construction and programming/Report writing : Methodology, Basic Sigmoid model in Stage 1/2 and advanced sigmoid model with small resolution in stage 3, Height investigation in stage 4

Reda Lahlou s192431

Model construction and programming/Report writing: Combining the sigmoïd model with polynoms, Comparison with basic models

31761 - Renewables in Electricity Markets
September 7, 2020

Contents

1 Introduction

In order to participate to the energy markets, a renewable energy producer needs to be aware of the quantity he can generates at each time unit. This paper deals with the issue of predicting the power generation of a wind farm based on a given set of past data available. The wind farm taken as reference is a farm in western Denmark, Horns Rev with a nominal capacity of 160MW. The aim of this study is to determine and develop a forecast model that will be able to predict the wind generation for a given period of time as precisely as possible, within the framework of a four-stages forecasting competition.
For every stage, past data of the wind farm and weather forecast from the European Centre for Medium-range Weather Forecast [1] are available. The past data are composed of the normalized power measurement $P(t)$ and the zonal & meridional components of the wind *forecast* at 10 and 100 meters above the ground level u and v for every time stamps t between the 1st of January 2015 and the most recent hour before prediction. Given the wind forecast in term of zonal and meridional components for a set period, the model to be build should be able to predict the normalized power generation for each time stamps t.
This paper introduced the methodology used in order to build a forecasting model based on sigmoid curves, and how this model was improved along the competition with different angular and wind speed resolution according to the intermediate and final results of each stages. These models were programmed in Python language on the software Jupyter. [3]

2 Methodology

2.1 Conversion of the wind coordinate system from Cartesian to polar

In every stage, two set of coordinates u and v for the forecast wind speed vector are given. One at 10 meters and one at 100 meters. These vectors are given in the Cartesian coordinate system: u is the zonal component of the wind forecast and v is the meridian one.
Speed and direction of the wind in the farm are more difficult to apprehend with this system, that is why a polar coordinate system would be more appropriate for interpretation since it will give the amplitude of the wind and its direction in the windmill.
The first step at each stage is to convert both wind vectors in the polar coordinate system.
For a wind vector (u, v) the speed module is :

$$s = \sqrt{u^2 + v^2} \tag{1}$$

and the orientation θ in the farm is :

$$\theta = tan_{ext}^{-1}(\frac{v}{u}) \tag{2}$$

where tan_{ext}^{-1} is the arctangent function extended to the whole angle circle.
Once the conversion has been made, the wind orientation is converted into degrees. From now on, every angle value will be given and discussed in degree.

2.2 Definition of the sigmoid

As shown later in this report, the function chosen to model the turbines power curve is a sigmoid. This curve has two plateaus in ∞ and $-\infty$ and a transitive curve in between from one plateau to the other. The sigmoid will be defined for the next experimentation as following :

$$f(x) = \frac{1}{1 + e^{-\frac{(x-r)}{s}}} \tag{3}$$

Where x is the wind speed and r and s are two parameters which modify the shape of the curve. r shifts the curve on the horizontal axis, acting as a delay ($f(r) = 0.5$ representing the speed at which the power worth half of the maximum power). s modifies the speed at which the curve will go from one plateau to the other.

2.3 Model selection and validation

During the four weeks of this challenge, many attempts have been tested in order to improve the model. In such a context, one needs to define a methodology for model training, selection and validation. This is the purpose of the following paragraph.

First some criteria for the fitting quality of a model were defined. If the real power value at time t is written $P(t)$ and the estimated one is $\widehat{P(t)}$, then the root mean square error $RMSE_i$ is defined for a lead time i :

$$RMSE_i = \sqrt{\frac{1}{T}\sum_{t=0}^{T}(P(t) - \widehat{P(t)})^2} \tag{4}$$

A final RMSE is also defined by the average of the RMSE for all lead times.
In the same way, the mean absolute error MAE for a lead time i is given by :

$$MAE_i = \frac{1}{T}\sum_{t}^{T}|P(t) - \widehat{P(t)}| \tag{5}$$

And finally the BIAS indicator is :

$$BIAS_i = \frac{1}{T}\sum_{t}^{T}(P(t) - \widehat{P(t)}) \tag{6}$$

The final fitting criteria will be the mean value of the three indicators for the 24 lead times.

It exists criteria that takes into account both the fitting quality and the complexity of the problem like the akaike criterion of the Bayesian information criterion. However, here the complexity of the problem and the number of parameters will only be discussed in a qualitative way since the quality of the fit is the main goal of the challenge as it is the only aspect that rank groups.

Finally, every model created was trained on data from 01/01/2015 01:00 to 30/06/2017 23:00. This represents about 2/3 of the training data. Then the models were tested on estimation of the data left, i.e from 01/07/2017 01:00 to the end. Some models were run on other period of the same size when a confirmation of the tendency was needed. As more data were made available after each stage, the number of validation time units increased from one stage to another.

3 Model development through the stages

3.1 Stages 1 & 2 : Data analysis and Simple model

3.1.1 Digging into data

In order to determine the type of model adapted to the power forecasting of the plant studied, a short analysis of the data available was conducted.
As the wind speeds at 10m and 100m are given in term of geographical component u and v, the

module and argument of the forecast wind speeds at each moment were calculated according to Equation (1) introduced previously. The wind power actually produced by the farm was drawn as a function of both forecast argument and module for every available past hours :

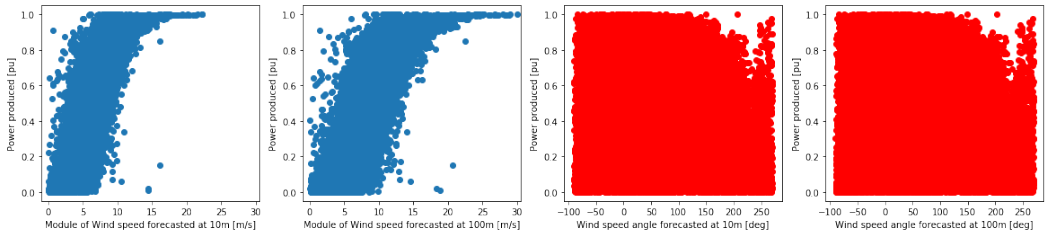

Figure 1: Power production VS Wind speed module and argument (Forecast) between 2015 and 2017.

Figure (1) gives information on the relation between the amplitude of the wind speed and the power produced by the plant. Indeed, at both heights the plot gives the shape of a sigmoïd curve between 0 and 1, for a range of speed between 0 and 30 m/s. On the other hand, the argument curves do not show any particular relation between argument and power, even if it seems like high values of power have never been reached for angles close to 230 degree.

Intuitively, one may think that the angle of the wind blowing on a farm will affect the turbine behaviour in term of power production. That is why the relation between wind speed amplitude and power was studied for different angular values (between 0 and 360 degrees) as shown in Figure (2).

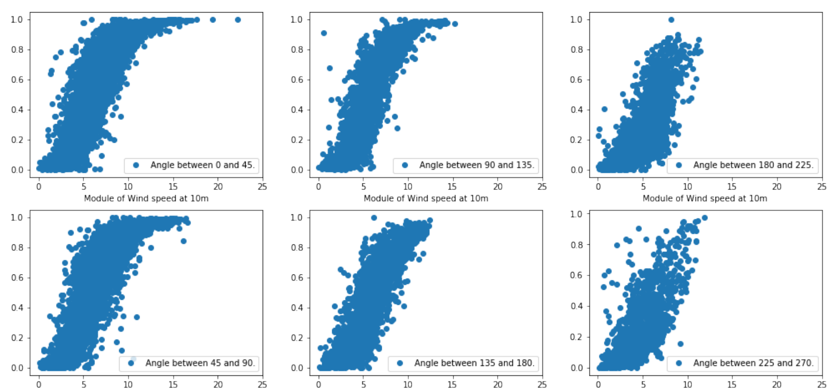

Figure 2: Power production VS Wind speed module for different angle between 2015 and 2017.

Here it appears that the curves do not behave in the same way depending on the predicted angle of attack of the wind. Thus, it might be interesting to consider both wind speed amplitude and angle to determine the power to be produced.

Finally, the data available mention the wind characteristics at 10m and 100m. According to the literature [2], the turbines studied here stand at 70m. The relation between wind speed properties and height has been studied to determine how the two set of data for two height can be combined or not, as depicted in the Figure (3). For both angle and amplitude, a linear relation with height can be estimated as first approximation :

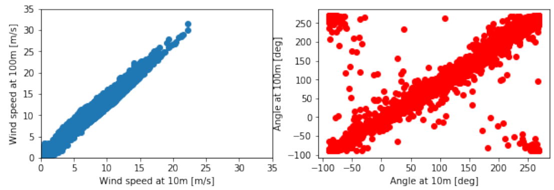

Figure 3: Data at 10m VS Data at 100m, for time units between 2015 and 2017.

Considering this first observations, some clues for the model have been found. Firstly, the power production can be seen as a function of wind amplitude through a sigmoid curve. Secondly, this function depends on the angular properties of the wind at a certain level. Thirdly, the wind properties for a given height can be seen as linear function of the height value between 10 and 100 meter. As the turbines rotor stand at 70m, it has been decided to estimate the wind properties at $H = 70$ m for every time units t with a linear relation for both amplitude and angle.

3.1.2 Sigmoïd approximation

As a first model, a sigmoid curve with two r and s parameters (see Equation (3)) is determined with the training data to estimate the power of the farm as a function of the wind speed at 70m. This choice of sigmoid is comforted by the fact that power curves of wind turbines often have a sigmoid shape.
The least square regression gives an r value of 7.691 and an s value of 2.316. A first overview of how the sigmoid fits the data is given on Figure (4) where every data point is shown and where the red line represents the estimated sigmoid model.

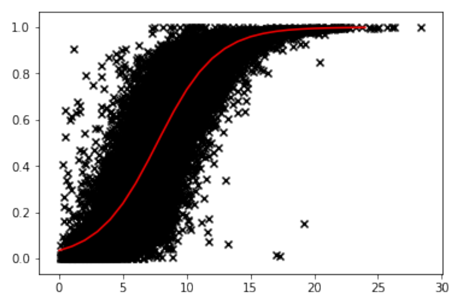

Figure 4: Model on Stage 1 & 2

3.1.3 Comparison with basic models

The performance of the model was assessed by comparing it to 2 benchmarks approaches:

- The **persistence model**, where the forecasted power at time t is defined as the last measurement observed: $\hat{P}_{t+k|t} = P_t \quad \forall k \geq t$
- The **climatology model**, where the forecasted power at time t is defined as the average of the measurements up to time t: $\hat{P}_{t+k|t} = \bar{P}_t \quad \forall k \geq t$

Figure (5) shows the average RMSE calculated for each lead time using the train and test data defined in section 2.3. The sigmoïd model outperforms the benchmark approaches by a large margin.

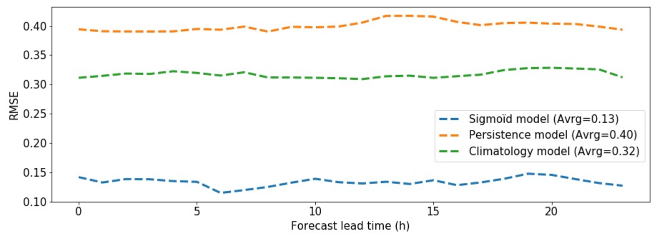

Figure 5: Comparison of average RMSE per lead-time

3.1.4 Analysis of the results

The following Figure (6) presents the performances of the model used in term of Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Relative Error (BIAS) per lead time for the prediction of January 2018:

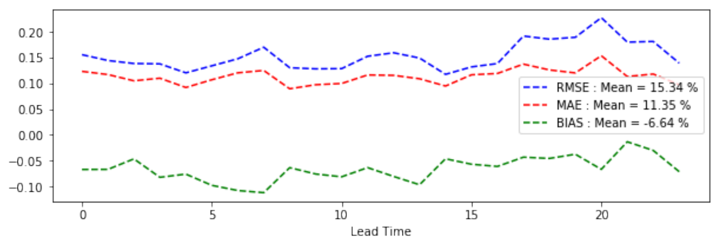

Figure 6: Error indicators for Stage 1 & 2, per lead time of forecast.

The BIAS curve shows that the error is always negative on average, meaning that the model tends to predict at higher power. No dependency between lead time and error can be seen from this figure, as both curves look flat. Indeed, the model is based on wind forecast and actual generation. Thus, the forecast error in term of wind speed cannot be seen through the data. The average RMSE is 15,34%. According to standard performance of wind power prediction, the model error can be reduced.
Taking a look at the absolute error compared to the wind speed, a tendency appears:

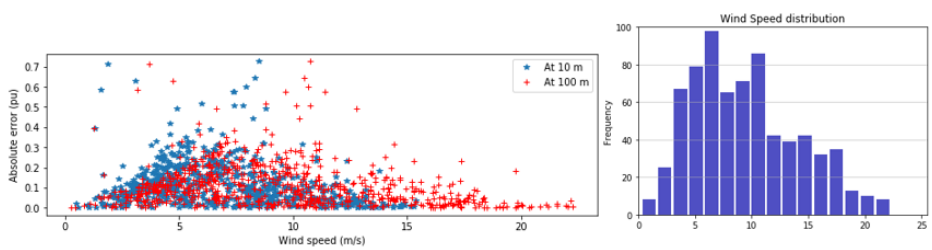

Figure 7: Absolute error VS Wind speed amplitude (forecast) in January 2018 & Wind Speed distribution at 100m.

The major part of the errors occurred for wind speeds between 5 and 10 m/s. These speeds correspond to the linear part of the sigmoid model where the data curve is the thickest. The model might then be improved in using more precise sigmoid curves trained with less data, depending an the angle of the wind for example. This tendency is confirmed by Figure (8) where it can be seen that prediction and generation are close for production close to the rated power (ie. 1pu) and more random for small power. Moreover, the model cannot predict high variation in the generation, which can be due to errors in the wind forecast.

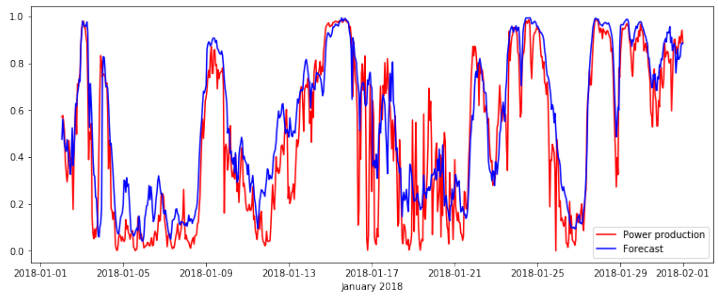

Figure 8: Predicted VS Real power production for Stage 1 & 2.

3.2 Stage 3 : Model improvements

Considering the previous observations, some studies were conducted in order to improve the model.

3.2.1 Several sub-models

As an improvement for this stage, different models were fitted on the data based on the wind orientation. This idea comes from the fact that windmills can not convert the same amount of power for a wind come from behind or aside. This is a building limitation of the windmills. Hence, it has been decided to separate the data into several intervals of wind orientation and to fit one model per interval to see if it gives better results.

In a first assumption, it has been decided to defined a regular division of the angles such that all intervals would have the same angle length. If an angle resolution ϕ is used, then the forecast model would consist in $\frac{360}{\phi}$ sub-models.
Different values of resolution ϕ have been tested. Of all tested resolution, the one retained is a resolution of 10, giving 36 sub-models. It has been seen that small resolution performed much better than high ones. Some resolutions smaller than 10 performed even better but it has been decided to keep 10 because the complexity added by decreasing the resolution was only making negligible improvements while increasing significantly the running time process.
In the process, one difficulty has been faced: the training data does not have an uniform distribution on angle. This can be illustrated with Figure 9 which shows every observation on 36 different graphs, each one representing an interval of 10 degree. Observations with a wind orientation between 150 and 270 are more rare and observations with such an angle are only for small wind speeds.
An idea to fix this issue was to separate the angles into intervals with a moving resolution so that all intervals would have the same number of observation points. However, this idea has been dropped because the new model brings only a negligible improvement in terms of prediction errors.

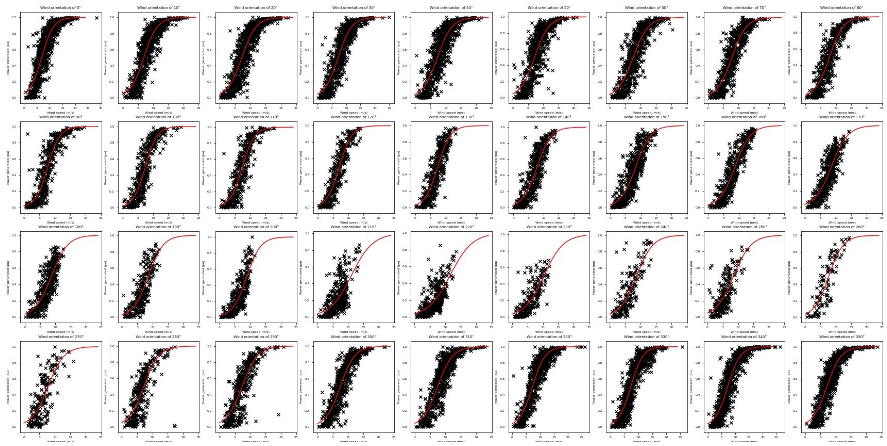

Figure 9: Distribution of the training data based on the wind orientation (interval of 10°)

3.2.2 Resolution per Lead Time

In parallel with the analysis of the effect of angular resolution, the effect of a resolution per lead time was studied, first for a angular resolution of 360 degree and then with the model introduced above.

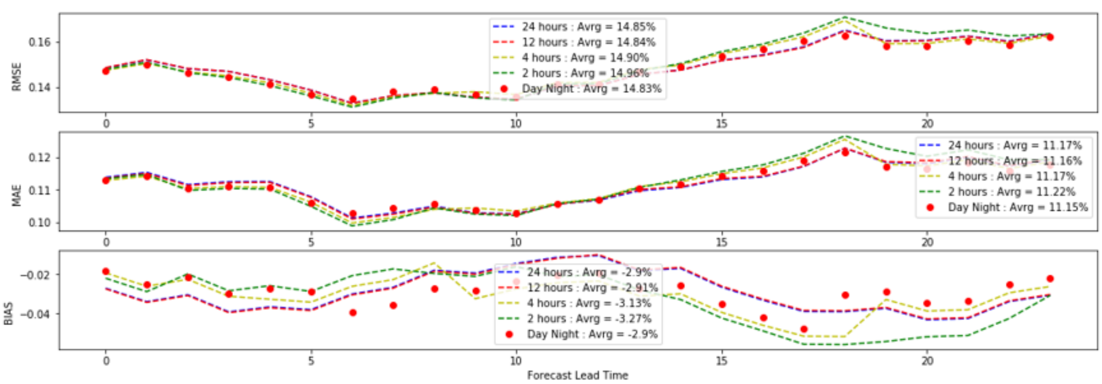

Figure 10: Error indicators per lead time of forecast for several Model resolution (Average of three tests).

This study reveals that, with a 360 angular resolution, it is more profitable to consider two models based on lead time hours of the wind forecast, as shown by the error indicators per hour and on average. However, the improvement of the performance is relatively low. Nevertheless, it is interesting to notice that in every case, two tendencies appears for day and night hours. Based on this observation, a new resolution of 12 hour was set, but sorting the lead time per "day hour" and not per number. The best choice of day/night hours was found to be 7 to 18 for day and 19 to 6 for night. This resolution offers better performances of forecast, even if again the gain is low.

However, when including this conclusion in the model with a more precise angular resolution, it has been seen that the model reliability was decreased as the root mean square error was increased from 13.06% to 13.47%. It has been supposed that the lead time influence observed was coming from the error made by the wide angular resolution, since the influence of the angular resolution has a higher impact on the forecast reliability.

3.2.3 Combining the sigmoid model with polynoms

The fits provided by the sigmoids on the train data (see Figure (9)) do not appear to follow properly the trends of the observations for low values of wind speed. To confirm this visual impression, the average RMSE in the test data was plotted as a function of wind speed on Figure (11):

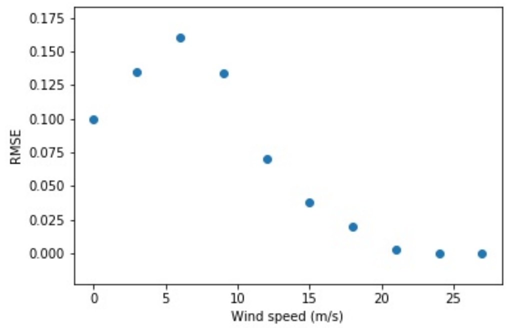

Figure 11: Average RMSE per wind speed value

In order to improve the model in low wind-speed regions, the possibility to use polynomial fits has been studied. This could help follow more accurately the beginning of the power curve. Different values for the wind speed limit have been tested (before this value, a polynom is fit, and the sigmoid model is used after). Polynoms of orders 1 and 2 have been tested - for higher order polynoms, evidence of over-fitting has appeared. The average RMSEs corresponding to these tests are shown on Table (1)

Polynoms degree\Wind speed limit	3	4	5	6
1	13.052	13.031	13.041	13.075
2	13.069	13.051	13.032	13.069

Table 1: Average RMSE combining polynomial and sigmoïd models

The lowest RMSE is found when fitting first-order polynoms on wind speed below 4m/s. The corresponding RMSE is 13.031, which is a slight improvement over the reference RMSE of 13.056 (using only the sigmoïd model). It has thus been chosen to include this combination of model for future predictions. Figure (15) (see in Appendix) shows the polynomial fits on the test data.

3.2.4 Analysis of the results

Figure (12) shows the indicators of the error made during the prediction of power generation in February 2018 using the improved model. Similarly to the previous stage, the curves are flat and do not present a particular dependency between error and forecast lead time. The relative error is still negative on average, but the global mean error is now of -4.3%, instead of -6,7%.

Figure 12: Forecast error indicators per lead time for February 2018 (Stage 3).

The average RMSE is now 11.31% against 15.34% for the previous stage. One may think that the model used is better to predict power production with the improvement introduced previously. Nevertheless, it is interesting to notice that the real power production was smoother in February, as shown is the Figure (13). The wind farm faced less high variation in the generation, what may have increase the forecast reliability in this case.

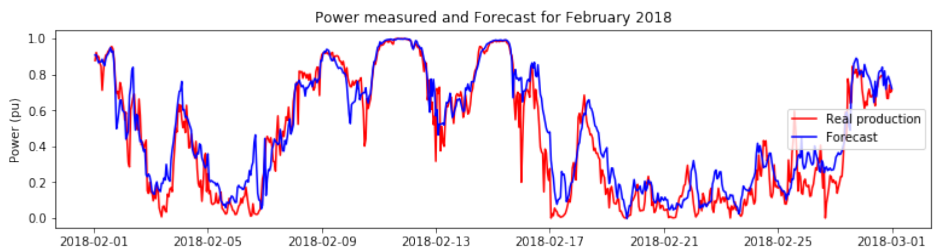

Figure 13: Predicted VS Real power production for Stage 3.

Here again, it can be seen that the major part of the error occurs for small generation, and so small wind speed. However, taking a look at the relation between speed amplitude and forecast error, it can be seen that for a wider range of wind speed (see distribution on Figure (14)) errors are smaller. The tendency for wind speed between 5 and 10 m/s remains but is less pronounced. The new model effects can then be approved.

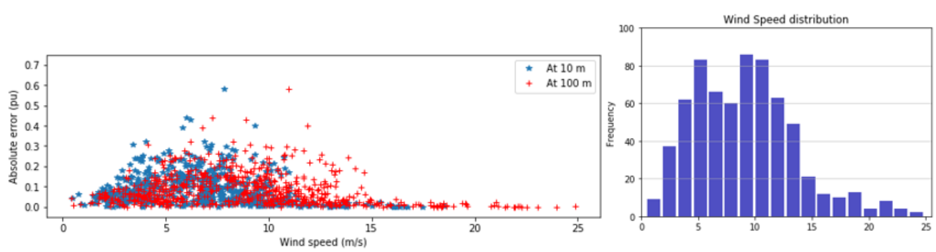

Figure 14: Absolute error VS Wind speed amplitude (forecast) in February 2018 & Wind Speed distribution at 100m

3.3 Stage 4 : Final model

From the first stage, the assumption has been made that the wind speed between 10m and 100m will be a linear interpolation between these two values and that the power selected will be based on the value of the wind at 70m. Here, this assumption is questioned. A logarithmic interpolation of the wind speed will be tried and all heights between 30 m and 100m will be tested to forecast the power based on the wind speed.
First with the linear model, an improvement in the RMSE and MAE is made by taking the

wind speed at lower values than the one at 70m. The best value is met at 50m, which allows to decrease the RMSE of the model by 0.3 and to achieve a RMSE below 13 on the test period for the very first time of the modelling process.
Then all height between 30m and 100m have also been tested with a logarithmic interpolation of the wind speed and none of them performs better than the linear interpolation. As an illustration, the better result achieved by the logarithmic model is at the height of 40m and gives an RMSE of 13.207 on the testing period while the linear model at height 50m gives an RMSE of 12.959.
The linear interpolation at height 50m is kept in the final model of stage 4.

4 Conclusion

The aim of this project was to construct a model to predict the hourly power output of a wind farm based on wind forecasts. The forecasting method developed relies primarily on the relation between the power produced by a wind turbine and the wind speed (called *power curve*). A sigmoïd function was used to estimate this power curve, with parameters determined by a least-squares estimation. Several attempts to enhance the model performance have been conducted, using RMSE as a criteria and splitting the past observations into train and test data. The final model consists in 36 sub-models for different wind direction, using a constant interval of 10$\circ$. A first-order polynomial fit was carried out for wind speeds below 4m/s, while the sigmoïd fit was employed for higher wind speeds. The regressor variable is the predicted wind speed at 50m height, found by performing a linear interpolation between values at 10m and 100m.

This model yielded satisfying results, with RMSE values down to 11.3% for predictions at stage 3. Significant improvements could potentially be reached if more variables were integrated into the model, such as the percentage of wind turbines under maintenance or the amount of curtailment performed. Plus, one should keep in mind that the data for wind speed are predictions. This limits the possibility of improving the model, as some significant unpredictable errors are bound to occur when predictions for wind are far from the real values.

5 Appendix

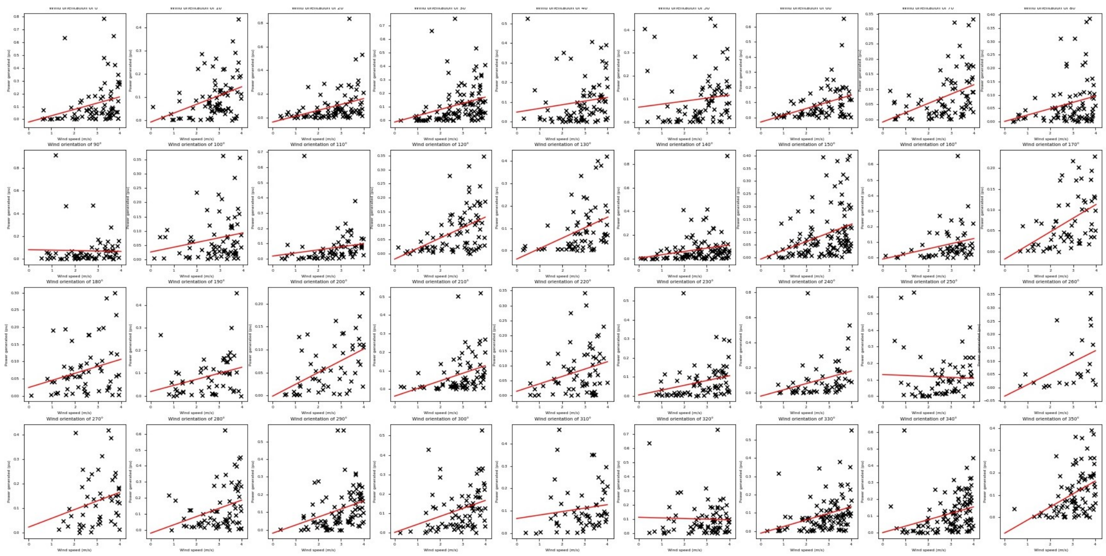

Figure 15: Polynomial fits for low wind speed values

References

[1] ”European Centre for Medium-range Weather Forecast” - www.ecmwf.int

[2] ”Horns Rev Offshore Wind Farm” - Wikipedia

[3] Jupyter software - Jupyter.org