Improvement of the standard Site Calibration for wind turbines

Irma García Galadí
irma.galadi@tecnico.ulisboa.pt

Instituto Superior Técnico, Universidade de Lisboa, Portugal

December 2021

Abstract

Site Calibration is the previous step to Power Curve Verification for wind turbines on complex terrains. An uneven Site Calibration can affect the Measured Power Curve and increase the risk of compensation. Nowadays, Site Calibration can be performed by multi-binning methods or using CFD tools as described in the IEC 61400 – Part 12. This Master Thesis proposes a new Site Calibration methodology that consists of using the existing reference met masts data as the input of Machine Learning regression models to better predict the wind speed at the turbine location and hub height. Nine wind turbines at three different locations are the object of this study. Three Machine Learning techniques are implemented: linear modelling Polynomial regression considering both Ridge and Lasso regularization, Artificial Neural Networks and the decision tree-based Extreme Gradient Boosting techniques. The main outcome of this research is that Machine Learning models applied to Site Calibration are more accurate than the current IEC 2005 standards and improve the Measured Power Curve estimation. Extreme Gradient Boosting outperformed in an average difference of more than 30% for the RMSE wind speed error and around 29.3% in power by wind turbine compared to the IEC 2005 baseline. Universal models by wind farm still perform better than the standards. And finally, SHAP values explainability tool points out the most important variables: wind speeds and wind directions at different heights including Turbulence Intensity due to its non-linearity. Furthermore, the most important sensors are the anemometer at the hub height and the ultrasonic anemometer.
Keywords: Site Calibration, Wind turbines, Machine Learning, IEC standards, Power Curve Measurement campaigns.

1. Introduction

Wind power is one of the fastest-growing renewable energy technologies. However, a high degree of certainty in investments return is the key for the development and execution of renewable projects. In the wind industry, particular attention is paid to the Annual Energy Production (AEP) due to its influence on the revenues of a project. A major concern for the AEP is the wind turbine Power Curve (PC) which expresses the relationship between the wind speed at the hub height and the power output. In this sense, the most broadly accepted Power Performance Testing (PPT) for verifying the correct performance of wind turbines and thus assessing the Warranted Power Curve (WPC), is the Power Curve Measurement campaign. It consists of measuring the power output of the wind turbine after its commissioning, being the Measured Power Curve (MPC) the result of this verification. When it comes to complex terrains, the PPT is a two-step procedure that involves the 'Site Calibration' campaign prior to the PCM.

Site Calibration (SC) is the method used to enable Power Curve Verification in complex terrains. Obstacles and surface roughness may disturb the airflow between the position of the meteorological (met) mast used as a reference and the centre of the turbine rotor. The SC approach estimates the impact of such disturbance and its uncertainty. Fig. 1 shows the standard SC setup which consists in placing two met masts, one in the reference position also called Permanent Mast (PM) and the second in the turbine location, known as Temporary Mast (TM) before the machine is commissioned. During the study period, the met masts record the wind speeds at the same time in each location. Once the recording period ends, the analysis is performed, and the turbine is erected. The SC procedure allows estimating a function or functions that transform the wind speed from the PM into the wind speed at the turbine position.

Nowadays, there are many different methodologies for carrying out a SC procedure and it is an open field of research inside of the industry.

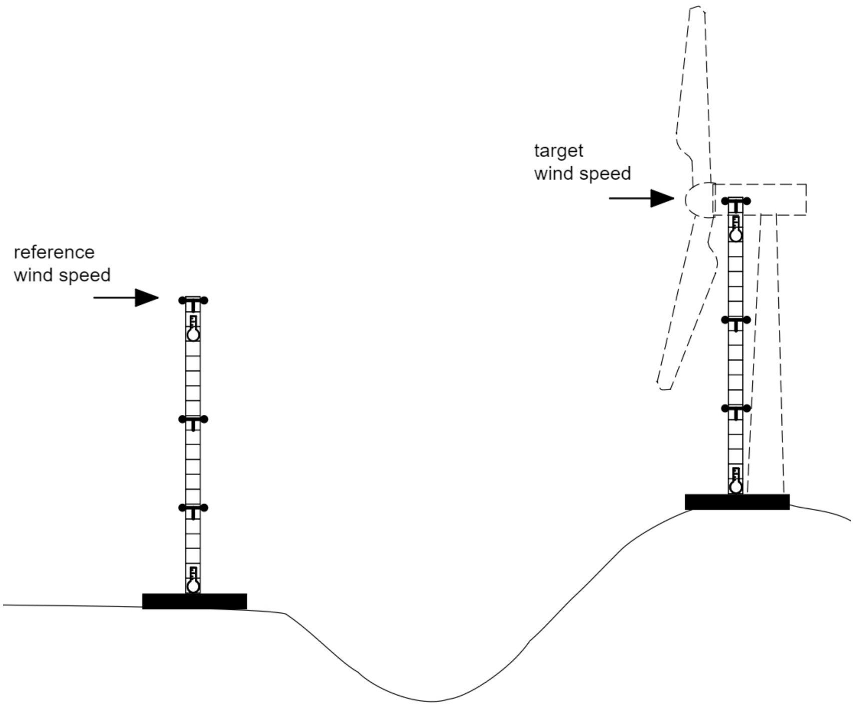

Figure 1: Standard Site Calibration setup

Thus, the main goal of this research have been to develop a reliable, efficient and non-expensive method for SC. The proposed methodology is consists of using PM data as the input to different Machine Learning (ML) models. The output of the modelling will be the predicted wind speed at the turbine location and hub height, given by the TM.

Other secondary objectives have been designing a practical implementation that may allow for a more generalized calibration of the terrain by wind farm and also identifying the most important sensors for SC. Accomplishing this latst objective may eliminate the need for installing certain sensors on the met masts, reducing the total cost of Power Curve Measurement campaigns.

2. Background

Currently, there are two main methodology groups for conducting a SC: the Annex C of IEC standards Part 12-1: 2005 [5] and 2017 [6] which uses a simple multi-binning methodology and the actual main alternative which is the Part 12-4 in the IEC 2017 standards. This is a physics-based SC methodology known as Numerical Site Calibration (NSC). Finally, advanced anemometry technology may be a future solution for measuring the wind speed at the turbine position and the hub height.

Regarding the IEC, the function or functions that convert the reference wind speed into the target wind speed is a table of Flow Correction Factors (FCFs) that depend either on the wind direction bin (at a minimum of 10 degrees bin), on the wind shear or on the wind speed itself. The FCFs are computed according to Eq. 1:

$$FCF_k = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{WS_{\text{target}}}{WS_{\text{reference}}}\right)_k \tag{1}$$

The FCFs are computed as the average wind speed ratio for the n observations and may depend on certain a k atmospheric variable depending on the IEC edition. IEC 2005 is the broadly accepted standard in the wind industry for which k corresponds only to wind direction bin. This is the version used as reference by this paper.

On the other hand, NSC is a three-dimensional flow analysis method based on Computational Fluid Dynamics (CFD) simulations to forecast the wind condition utilizing the correlations of wind characteristics between the reference site and the prospective wind turbine installation site [8]. Although this method might present some advantages, it is computational expensive, the uncertainty in AEP/performance is high and it is perceived as inconvenient due to a lack of knowledge from the industry point of view.

And finally, earlier in 2001, J.P. Verhoef and G.P. Leendertse [12] had already pointed out the necessity of exploring different ML regression techniques for SC procedures. However, despite that ANNs were recommended for solving the SC problem 20 years ago, this researcher could not find any report, article, or document regarding the application of any ML technique to a SC problem.

A SC procedure can have a significant impact on the final MPC which is the result of a Power Curve Measurement campaign. On a Turbine Supply Agreement between the parties, the PC provided by the manufacturer is the actual Warranted Power Curve (WPC). The validation of the WPC is done by simultaneously recording the wind speed at the PM and the wind turbine power output. Once the data is collected, it should be normalized to a reference air density according to Eqs. 2 and 3: [5]. The reference density may be a pre-defined nominal air density representative of the site.

$$V_n = V_m \left(\frac{\rho_m}{\rho_o}\right)^{\frac{1}{3}}, \tag{2}$$

Where V_n is the normalized wind speed, V_m is the measured wind speed, ρ_m is the measured air density and ρ_o is the reference air density.

$$P_n = P_m \left(\frac{\rho_m}{\rho_o}\right), \tag{3}$$

Where P_n is the normalized power output and P_m is the measured power output.

Once the data is normalized, the wind speed is binned and the average power for each wind speed bin is computed. The result of this computation is the MPC.

Usually, a warranty contract stipulates that the Measured AEP (MAEP) of the wind turbine shall be equal or greater than Warranted AEP (WAEP). The equations for AEP (Measured and Warranted) are:

$$MAEP = MPC \cdot WSD \tag{4}$$

$$WAEP = WPC \cdot WSD \cdot (1 - u_{\text{AEP}}) \tag{5}$$

Where u_{AEP} is the uncertainty in the AEP, which although it is usually formulated and computed in warranty contracts, is not addressed in this paper.

However, when it comes to complex terrains, the wind speed at the turbine location considered in the estimation of the MPC is computed as described in the IEC standards, as shown in Eq. 6:

$$WS_{\text{target}} = FCF_{WDbin} \cdot (WS_{\text{reference}})_{WDbin} \tag{6}$$

Therefore, SC plays an important role when estimating the MPC in complex terrains. Through the FCFs calculated during the IEC SC procedure, together with the reference wind speed measured during the Power Curve Measurement campaigns, the target wind speed is estimated. If an inaccurate target wind speed is estimated, an uneven SC can transfer a significant prediction error to the wind speed binning used for estimating the MPC. The risk of compensation on behalf of the turbine supplier increases significantly when the target wind speed is over-predicted for a wind speed since it will lead to a lower MPC. This increase is specially risky when the wind speed over-estimation is that to the wind speed corresponding to power below the rated power because this error is transferred to the MPC to the third power, 7:

$$P = \frac{1}{2} C_p \, \rho \, \pi \frac{D^2}{4} U^3, \tag{7}$$

Where C_p is the power coefficient, around 0.593. ρ is the air density in kg/m^3, D is the wind turbine rotor diameter in meters and U is the wind speed at the turbine height in m/s.

A method was developed by the Department of Power Curve Verification of Vestas to quantify the impact of an “uneven” SC procedure in energy terms. It consists in obtaining a ‘Site Calibration Power Curve’ (SC-PC) from the SC data. The SC-PC combines the predicted wind speed at the turbine location and the power derived from the target wind speed at the turbine location using the WPC as a reference [11].

Fig. 2 shows an example of this method for the wind turbine WTG14. The blue scatter on the x-axis corresponds to the predicted wind speed by the IEC method, while on the y-axis the power calculated based on the turbine mast location measured wind speed is shown and can be considered as the “expected power”. Also, the black dotted line represents the WPC provided by Vestas. While the red dotted line is the SC-PC which represents the SC error.

When the computed FCFs over-estimate the wind speed, the blue scatter points are shifted to the right and the red line SC-PC is forced downward. Thus, it can be stated that the standard SC error is transferred to the wind speed prediction binning of the MPC. Please note that for a SC error equal to zero, black (WPC) and red lines (SC-PC) should be equal.

Both Measured and Warranted AEP are computed using Eqs. 4 and 5 and their difference is used for measuring the error in energy terms through Eqs. 8 and 9.

$$AEP_{\text{diff}} = MAEP - WAEP \tag{8}$$

$$AEP_{\text{percentage}} = \frac{MAEP}{WAEP} \cdot 100 \tag{9}$$

For $AEP_{\text{diff}} < 0$ or $AEP_{\text{percentage}} < 100\%$, the SC procedure would increase the turbine supplier risk of compensation.

3. Proposed Methodology

The proposed methodology consists of using the wind and other meteorological variables measured

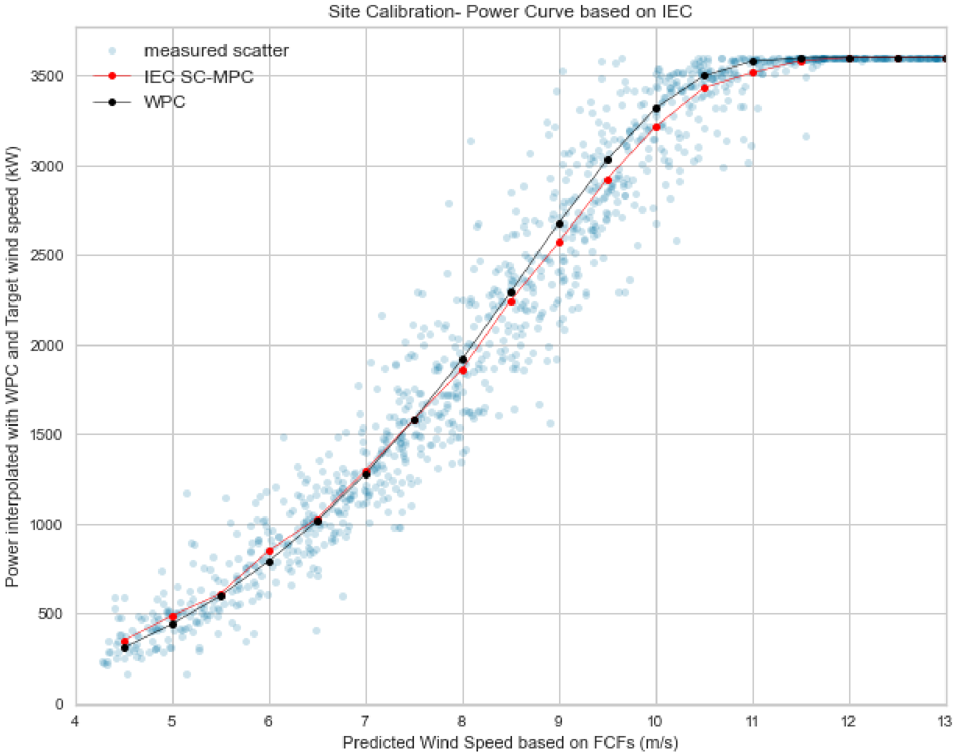

Figure 2: Site Calibration-Power Curve of IEC for WTG14

by the Wind Measurement Equipment (WME) on the PM was the input data for the ML models implemented.

3.1. Data

As shown in Fig. 3, the met mast is a steel tower on the top of which the WME is mounted. The digital Supervisory Control And Data Acquisition (SCADA) system which collects the data continuously at a sampling rate of 1 Hz (every second) for most of the signals. The pre-processed data is sampled in a base of 10-minute periods: mean, standard deviation, maximum and minimum. Each of these points in a dataset will be named "an observation" from now on.

In ML, each variable used as input for modelling is known as 'feature', so from now on, the meteorological variables used for modelling will be named as input features. The input features considered are 10 minute-averaged wind speed and wind direction at the hub, mid tip and lower tip heights, wind shear, turbulence intensity and wind veer. From the 3D anemometer mounted at the hub height, the wind speed horizonal and vertical components as well as for the wind direction. Also, relative humidity, temperature, pressure and air density were considered at two different heights. The formulas for the post-processed variables are as follows: Air Density is computed as shown in Eq. 10.

$$AD1 = \frac{1}{T1 + 273.15} \left[\frac{PR_{\text{hub}} 100}{R_{\text{o}}} - \frac{RH1}{100}\, p_{\text{vapor}} \left(\frac{1}{R_{\text{o}}} - \frac{1}{R_{\text{w}}} \right) \right], \tag{10}$$

- Where PR_{hub} is the pressure at the hub height [hPa], which may not correspond to the height at which the pressure sensor is set. Thus:

$$PR_{\text{hub}} = PR1 - \Delta z\, \Delta PR \tag{11}$$

  - Delta height [m], $\Delta z = z_{\text{hub}} - z_1$
  - Delta pressure [hPa], $\Delta PR = \left[PR_{\text{o}} - PR_{\text{o}}\, e^{\left(\frac{-9.8}{R_{\text{o}}(15+273.15)}\right)} \right] 0.01$
  - Standard Pressure, $PR_{\text{o}} = 101325$ Pa

- Dry air constant

$$R_{\text{o}} = 287.05\ J/(kgK) \tag{12}$$

- Water vapor gas constant

$$R_{\text{w}} = 461.5\ J/(kgK) \tag{13}$$

- Vapor pressure [Pa]

$$p_{\text{vapor}} = 0.0000205\, e^{(0.0631846\,(T1+273.15))} \tag{14}$$

Turbulence Intensity is computed only for wind speed at the hub height as shown in Eq. 15.

$$TI = \frac{WS1_{\text{std}}}{WS1_{\text{avg}}} \tag{15}$$

Wind Shear is computed for the wind profile described by the wind speed at lower tip level and that at the hub height.

$$WSH = \frac{\log\left(\frac{WS1}{WS4}\right)}{\log\left(\frac{z_1}{z_4}\right)} \tag{16}$$

Wind Veer is computed for the difference in wind direction for that at the hub height and at the lower tip level.

$$WVeer \quad = \quad \begin{cases} WD1 - WD4 - 360, \\ if\ WD1 - WD4 > 180 \\ WD1 - WD4 + 360, \\ if\ WD1 - WD4 < -180 \\ WD1 - WD4, \\ otherwise. \end{cases}$$

Also known as Vertical Wind Direction (WDVer), it is computed as the angle between the horizontal and the vertical wind speed measured by the ultrasonic anemometer.

$$WDVer \ = \ \frac{-180}{\pi}\ \arctan\left(\frac{WSVer}{WSHor}\right) \tag{17}$$

Regarding data mining, the implemented protocol consisted of three different steps: data pre-processing, data filtering and data conditioning. The first step involves transforming raw data into an understandable format ready to be explored. This first step included a data correction step for the Turbulence Intensity and the Inflow angle, a data cleaning process of the missing observations and a data quality control that consisted on detecting the faulty values for each variable defined by a range of quality and dropping the corresponding observations. After pre-processing, a data filtering process was applied. These filters correspond to the measurement sectors and the operational range of the wind turbine and the icing filter. Finally, a data conditioning step was put in place in order to prepare each dataset for the subsequent ML modelling. This step consisted of first, applying the “Hold-out” method which is splitting the raw dataset into two different datasets: the training set (70%) and the test set (30%). The ML models are trained and validated with the training set while are assessed with the test set, which is basically a complete unseen data series for the trained model. After splitting, each feature of the dataset was normalized.

The object of study are nine wind turbines located at three different sites in Australia. The first dataset includes data for two wind turbines (WTG14 and WTG15). For dataset2, three wind turbines are included in the study (T11, T17 and T22) and finally, for the third dataset four different wind turbines are analysed (WTG18, WTG20, WTG43, WTG46).

3.2. Modelling

In this subsection both the steps of the modelling and the different ML techniques implemented based on python code scripts.

3.2.1 Modelling steps

The four steps of any ML task are hyperparameter tuning, model training, generalization and model explainer.

Each algorithm requires a specific set of hyperparameters that need to be adjusted according to the task. Both Grid Search and Random Search are implemented for hyperparameter tuning through a as k-fold Cross-Validation [1]. Regularization may also be applied. It consists in constraining a model to make it simpler and reduce the risk of overfitting [1].

Training a model in ML is like solving an optimization problem. The model learns from the training set of inputs by adjusting the its parameters (also named coefficients) so that the difference between the predicted output and the target value is minimum. This difference is called ‘loss’ and it is expressed through the Mean Squared Error (MSE) as can be seen in Eq. 18.

$$\text{minimize}\ \frac{1}{n}\sum_{i=1}^{n}(y_i \ - \ \hat{y}_i)^2 \tag{18}$$

Where y_i is the target value and $\hat{y}_i$ is the predicted value for each of the n observations.

ML involves using an algorithm to learn and generalize from historical data to make predictions on new data. The variables that are entered into the model are called features and the output of the model is called a prediction.

The generalization consists in using the test set to assess how well the model can generalise to new and unseen data. Three different metrics are defined so that the models are comparable to both wind and power:

The Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i \ - \ \hat{y}_i|^2 \tag{19}$$

The Root Mean Squared Error (RMSE):

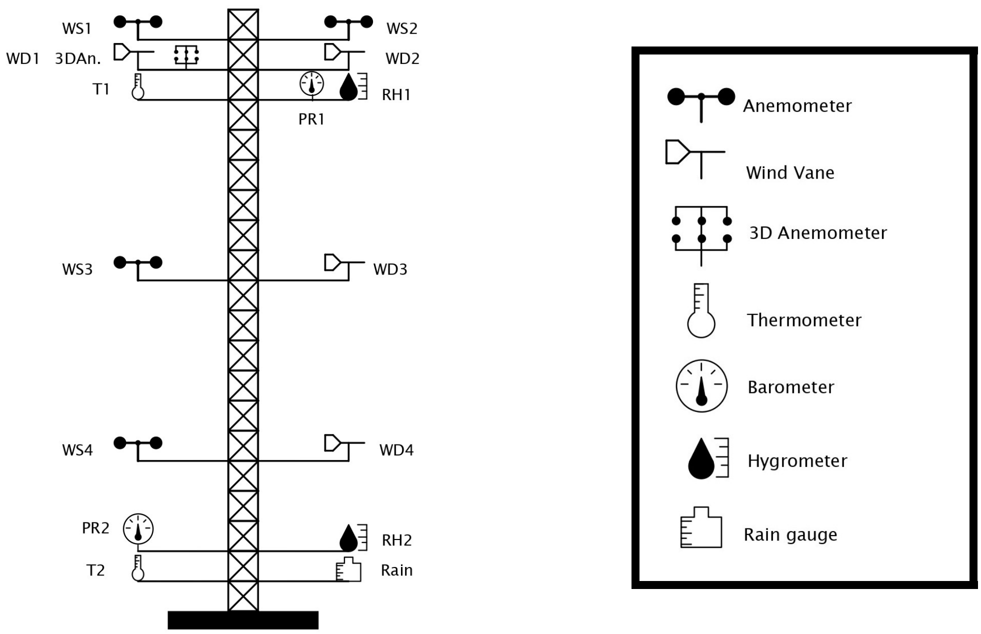

Figure 3: Meteorological mast with the Wind Measurement Equipment (Not at scale)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{20}$$

And the Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{21}$$

And finally, SHAP [9] is the state-of-the-art in ML models' explainability. SHAP values are a series of parameters that quantify the contribution of each feature to the prediction of a given complex model.

3.2.2 Modelling techniques

Three different ML techniques are implemented. Linear modelling Polynomial regression considering both Ridge and Lasso regularization, implemented on Scikit-Learn library [10], Artificial Neural Networks on Keras [3] and the decision tree-based Extreme Gradient Boosting techniques based on xgboost package [2].

A linear model in ML is any model that assumes a linear relationship between the input features and the output and those can be expressed explicitly. Two types of linear models are implemented: simple linear regression and polynomial linear regression. The equation of the simple linear model only considers the set of values x_{ij} for each of the p inputs, for each of the n observations, in their original form to predict the target value $\hat{y}_i$, through the optimization of the linear coefficients for each feature β_j and the independent coefficient β_0.

$$\hat{y}_i = \beta_0 + \sum_{j=1}^{p} (\beta_j + x_{ij}) \tag{22}$$

On the other hand, the polynomial regression inputs those features in a polynomial shape. For instance, if the problem had the features (a, b, c) the input polynomial features for degree equal to 2 would be $(1, a, b, c, a^2, b^2, c^2, ab, bc, ca)$ [4].

Ridge and LASSO regression are two regularization techniques that consist in adding a term to the error to reduce the value of the coefficients in the model. The level of penalty is controlled by the hyperparameter α. Eq. 24 presents the optimization for Ridge and Eq. 23 for LASSO.

$$\text{minimize } \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j + x_{ij} \right)^2 + \alpha \sum_{j=1}^{p} |\beta_j| \tag{23}$$

$$\text{minimize } \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j + x_{ij} \right)^2 + \alpha \sum_{j=1}^{p} \beta_j^2 \tag{24}$$

Artificial Neural Networks (ANN) is the second method implemented for solving the Site Calibration regression task. It is a type of non-linear ML modelling. An ANN is then an interconnected group of nodes, known as Artificial Neurons that through an iterative process can learn from a dataset to predict an output. The architecture of the ANN is shown in Fig. 4.

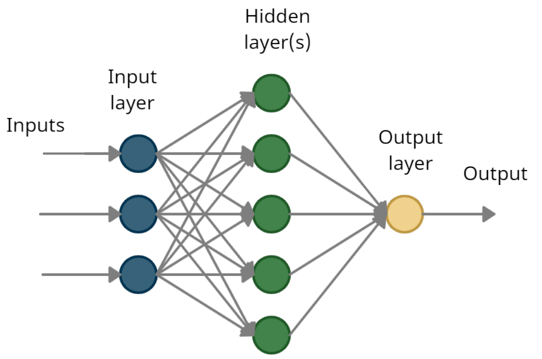

Figure 4: Basic Architecture of an Artificial Neural Network

The considered architecture is a fully connected network with one neuron at the output layer, since the SC task is to estimate one single target, the wind speed at the wind at the hub height on the turbine location. There 6 types of hyperparameters that are tuned through a Random Search for the ANN modelling. The number of hidden layers (up to 3) and the number of neurons per layer (up to 100), the learning rate, the regularization technique (L1, L2, Dropout or Early Stopping), the optimizers (SGD, Momentum, Nesterov, Adam, Nadam or RMSProp) and the activation function (ReLU, Leaky ReLU, eLU or SeLU).

The third and final ML implemented is the Extreme Gradient Boosting decision tree-based, which is an ensemble model. The idea behind ensemble modelling is that a single algorithm, on its own, might not be able to capture all the relations in a given dataset. However, a group of algorithms trained with different parts of a dataset might be able. Boosting trains models sequentially, each new model is trained to correct the errors of the previous ones.

On the other hand, as can be seen in Fig. 5, decision trees apply a top-down approach to data so that for given a dataset, the algorithm splits each region in a way that makes most training observations as close as possible to that predicted value. Up to 8 different hyperparameters are tunned for the XGB molldeing through an stepwise Grid Search as Aarshay Jain describes [7]. Among these hyperparameters, the most relevant are the number of estimators, the learning rate and the maximum depth of each tree.

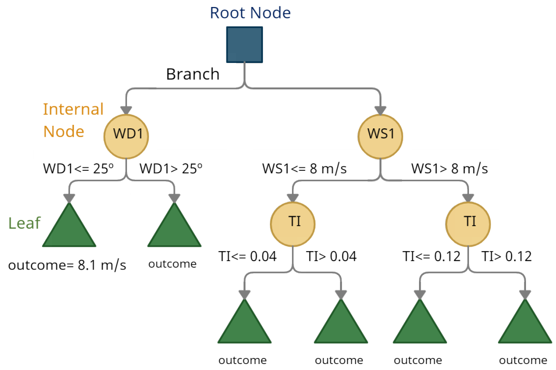

Figure 5: Basic Architecture of a Decision Tree

4. Results & discussion

The main goal of the present paper is to prove the validity of a data-driven model to perform a SC for wind turbines. Table 1 summarizes the average MAPE error by ML technique compared to the IEC 2005 standards. The main outcome from this comparison is that in all cases, ML models present a lower error than the standards. The error reduction by wind turbine for wind speed was from 0.7% up to 5.4% while for the power output, from 1.5% up to 15.6%.

Table 1: Average MAPE error comparison by ML technique

Average MAPE (%)		wind speed	power output
IEC		6.87%	15.61%
Linear	regular	5.33%	11.73%
	general	5.71%	14.90%
ANN	regular	5.12%	11.29%
	general	5.59%	15.10%
XGB	regular	4.77%	10.59%
	general	5.44%	13.60%

For individual models, the ML technique with better performance was XGB followed by ANN, being the linear models the less effective in predicting the wind speed. When looking at the Universal model by wind farm, although not being as accurate as individual models by wind turbine, they still perform better than the IEC standards.

From the wind industry perspective, the main concern raised is to find a SC that accurately predicts the wind speed at the hub height and at the turbine location and, more important than that, a model that correctly estimates the MPC and thus the MAEP.

Fig. 6 presents the SC-PC for wind turbine WTG14. The MPC based on XGB is closer to the expected WPC than the IEC curve, especially

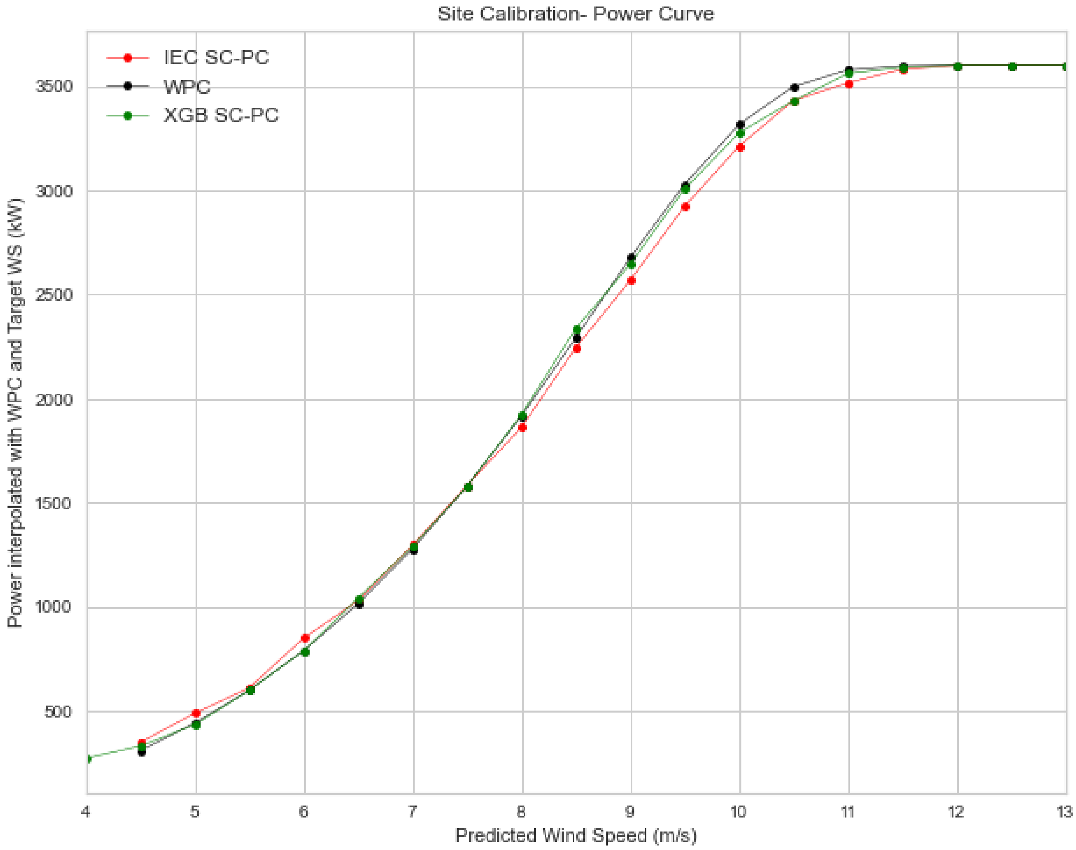

Figure 6: Site Calibration-Power Curve comparison for WTG14

around the elbow, but also on the lowest part of the curve. Regarding the MAEP, the XGB technique considerably improves the MAEP estimation compared to the IEC method for all cases with the exception of wind turbine WTG18, for which the difference with the expected AEP is similar. The average improvement is around 62% of the AEP baseline and in the best case up to a 89.7% improvement for wind turbine WTG14.

Feature importance depends mainly on the ML technique and, but also on the site. Fig. 7 thereupon shows the average feature importance by ML technique for all sites.

As seen in Fig. 7, for all three models, the two most important features are the wind speed at the hub height, as expected, but also its horizontal component captured by the ultrasonic or 3D anemometer. Moreover, WS3 and WS4 meteorological variables are ranked to be around the same position for all three models.

While linear models are much less selective, thus these models tend to give importance to more variables, non-linear models can better select the most important variables, especially XGB, which is very consistent regarding feature importance among the different wind turbines.

From all wind variables, the case of TI is worth to be mentioned. TI has a highly non-linear relationship with the target, that is why it is at the end of the ranking for the linear models while it is at the top for non-linear models, especially for XGB which TI can be found in the third place.

Although the SHAP values tool implemented provides the importance by feature, from the wind industry insights learned, the importance is required to be measured by sensor. Thus, one simple methodology has been developed to transform feature importance to sensor importance using Eq. 25:

$$importance_{\text{sensor}} = \sum_{i=1}^{n} \left(\frac{importance_{\text{feature}}}{\sum sensors} \right)_i \tag{25}$$

Fig. 8 shows the average sensor importance by ML technique estimated considering the necessary signals for computing the different features.

As can be seen in Fig. 8 the top 5 most important sensors for SC are exclusively wind speed and wind direction related and the least important in all cases is the Rain gauge. Looking at this ranking in detail, it is concluded that anemometer at the hub height and the 3D anemometer are a must in SC performed through ML techniques, while Temperature sensors and Relative Humidity and Pres-

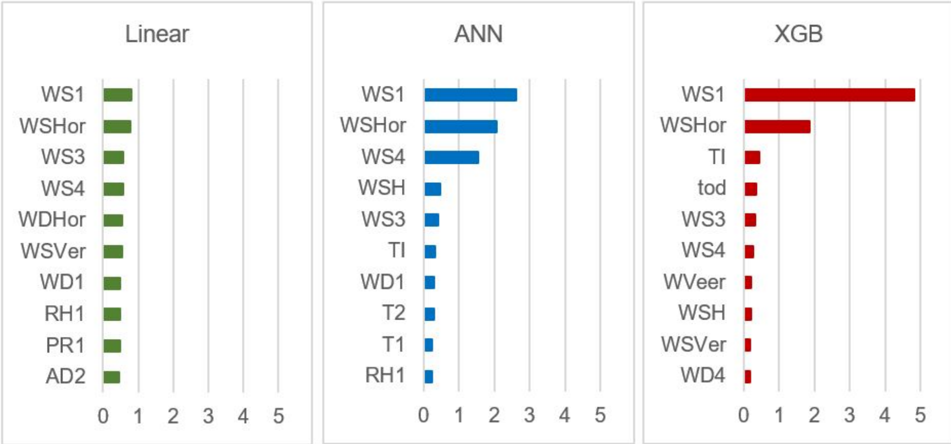

Figure 7: Average feature importance Top 10 Ranking by ML technique

sure sensors at the lowest tip level, far from the hub height, could be considered in the droplist.

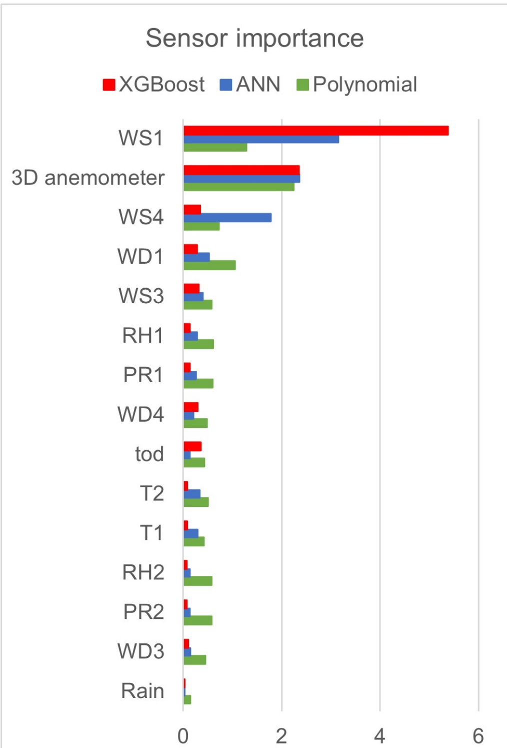

Figure 8: Sensor importance averaged for all sites by ML technique

5. Conclusions

As seen in Fig. 2, it has been proved that the IEC 2005 SC standard method systematically overestimates the wind speed around the rated power. The two main consequences are a significant decrease in the MPC at its elbow and in some cases, a reduction in the MAEP by default. Moreover, the wind speed is underpredicted on the lowest part of the Power Curve and thus, the MPC tends to increase. Therefore, the inaccuracy of wind speed prediction for SC procedures in some cases increases the risk of compensation on behalf of the turbine supplier.

ML supervised learning tools are more accurate than the IEC in predicting the wind speed at the hub height and wind turbine location when performing a SC procedure. In most cases, the systematic errors introduced in the MPC by the IEC standard method are corrected by the ML models implemented, especially XGB. And consequently, ML models are also more accurate in estimating the MAEP. This main conclusion can lead to other two relevant outcomes namely the assurance of development of new renewable projects as well as the integration of wind farms into the network.

First, regarding the main motivation of this Master Thesis, ML models applied to SC can potentially reduce the risk of compensation on behalf of the turbine supplier due to wind turbine underperformance. Moreover, it also can ensure the certainty of the investment return and the profitability for the wind farm owner which is the key for the development and execution of renewable projects.

On the other hand, unlike conventional power plants, wind farm power production is entirely dependent on environmental conditions at each wind turbine location. However, wind speed at the hub height and turbine location not only depends on an intermittent and variable wind resource at the reference location but also on complex non-linear atmospheric interactions. This reference meteorological masts data is later used for power production prediction for the wind farm daily operation. Thus, the increase in the accuracy of the power prediction thanks to a more accurate SC procedure is recognized as a major contribution to reliable large-scale wind power integration.

For all of the above, it is recommended to abandon the multi-binning linear regression methods and to adopt the multivariate non-linear regression models. Especially XGB modelling due to its outperforming results and its ability to accurately select the optimal required features. It is also recommended to include other meteorological variables rather than the wind speed and wind direction at the hub height, namely other wind-related variables at the middle tip and lower tip levels, the Turbulence Intensity and to consider the installation of ultrasonic anemometers on the PM.

Model explainability tools for ML models, such as SHAP values, can increase the transparency of the models which can be translated to increasing acceptance among the wind industry members. Moreover, model explainers applied to SC can also be useful for optimizing the budget of Power Curve Measurement campaigns by identifying the most relevant WME devices.

Finally, it is concluded that universal ML models by farm can perfectly be an option for pursuing a SC. Although universal models have not proved to be as accurate as regular individual models by wind turbine, they still outperform IEC standards. 'Universal models' may be seen as more convenient in terms of complexity.

Acknowledgements
I would like to express my great appreciation to Andreas Schmitz and Fernando de Freitas for their valuable and constructive suggestions during the planning and development of this research work. Their willingness to give their time so generously has been very much appreciated.

I also would like to express my deep gratitude to Prof. Ricardo Pereira, my thesis supervisor, for his patient guidance, enthusiastic encouragement, and also his insight, always support and sharing of knowledge that has made this Thesis possible.

I would like to thank my family for their encouragement and caring over all these years and for always being there for me through thick and thin.

Last but not least, to all my friends and colleagues that helped me grow as a person and were always there for me during the good and bad times.

To each and every one of you – Thank you.

References

[1] G. Aurélien. *Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow*. O'REILLY, 2 edition, sep 2019. powered by Jupyter.

[2] T. Chen and C. Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, NY, USA, 2016. ACM.

[3] F. Chollet et al. Keras. 2015.

[4] M. J. Grover and M. Maldonado. Polynomial Features and Regularization Demo - Part 3 - Regression with Regularization Techniques: Ridge, LASSO, and Elastic Net.

[5] International Electrotechnical Commission (IEC). *IEC 61400 – Part 12-1: Power Performance measurements of electricity producing wind turbines*, 2005.

[6] International Electrotechnical Commission (IEC). *IEC 61400 – Part 12-1: Power Performance measurements of electricity producing wind turbines*, 2.0 edition, 2017.

[7] A. Jain. XGBoost Parameters | XGBoost Parameter Tuning, Mar. 2016.

[8] J.-H. Jeong and K. Ha. Evaluation of wind flow characteristics by rans-based numerical site calibration (nsc) method with met-tower measurements and its application to a complex terrain. *Energies*, 13:5121, 10 2020.

[9] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc., 2017.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

[11] P. C. V. Team. Theoretical power performance estimation based on site calibration data. Internal technical report, Vestas, jul 2020.

[12] J. Verhoef and G. Leendertse. Identification of variables for site calibration and power curve assessment in complex terrain. task 8: A literature survey on theory and practice of parameter identification, specification and estimation (ise) techniques. ECN 7.4062 project report JOR3-CT98-0257, Energy research Centre of the Netherlands, apr 2001. Project Name: SiteParIden.