Wind Turbine Systematic Yaw Error: Operation Data Analysis Techniques for Detecting It and Assessing Its Performance Impact

Davide Astolfi [1,*]**, Francesco Castellani** [1] **, Matteo Becchetti** [1] **and Andrea Lombardi** [2] **and Ludovico Terzi** [2]

[1] Department of Engineering, University of Perugia, Via G. Duranti, 93-06125 Perugia, Italy; francesco.castellani@unipg.it (F.C.); matteo.becchetti@unipg.it (M.B.)

[2] Renvico srl, Via San Gregorio 34, 20124 Milano, Italy; andrea.lombardi@renvico.it (A.L.); ludovico.terzi@renvico.it (L.T.)

* Correspondence: davide.astolfi@unipg.it; Tel.: +39-075-585-3709

Abstract: The widespread availability of wind turbine operation data has considerably boosted the research and the applications for wind turbine monitoring. It is well established that a systematic misalignment of the wind turbine nacelle with respect to the wind direction has a remarkable impact in terms of down-performance, because the extracted power is in first approximation proportional to the cosine cube of the yaw angle. Nevertheless, due to the fact that in the wind farm practice the wind field facing the rotor is estimated through anemometers placed behind the rotor, it is challenging to robustly detect systematic yaw errors without the use of additional upwind sensory systems. Nevertheless, this objective is valuable because it involves the use of data that are available to wind farm practitioners at zero cost. On these grounds, the present work is a two-steps test case discussion. At first, a new method for systematic yaw error detection through operation data analysis is presented and is applied for individuating a misaligned multi-MW wind turbine. After the yaw error correction on the test case wind turbine, operation data of the whole wind farm are employed for an innovative assessment method of the performance improvement at the target wind turbine. The other wind turbines in the farm are employed as references and their operation data are used as input for a multivariate Kernel regression whose target is the power of the wind turbine of interest. Training the model with pre-correction data and validating on post-correction data, it is estimated that a systematic yaw error of 4° affects the performance up to the order of the 1.5% of the Annual Energy Production.

Keywords: wind energy; wind turbines; yaw system; performance monitoring; multivariate regression; operation data analysis

1. Introduction

Horizontal-axis wind turbines (HAWTs) represent one of the most promising renewable energy technologies, especially as the energy density increases with the recent developments in rotor size [1].

It is fundamental for improving the energy yield up to some percent of the annual energy production (AEP) that the wind turbine control is robust and that the blade aerodynamics are appropriately optimized [2]. A robust wind turbine control means also that the wind turbine should operate as much time as possible with the rotor plane facing perpendicularly the wind flow: the yaw angle has therefore a theoretical target of $0°$. In practice, this is not possible because of the mismatch between the time scale of the meandering wind direction and the large inertia of the nacelle, to be counteracted through the yaw motors.

Furthermore, the impact of yaw angle on the energy yield is remarkable: a theoretical estimation [3] is that the extracted power is proportional to the cosine cube of the yaw angle. In [4], the effect of yaw error depending on the operation regime has been investigated through a wind turbine model, yaw error model and the equivalent wind speed model that includes the wind shear and tower shadow effects: the conclusion is that the cosine cube law is simplistic and yaw error can remarkably affect the power output; for example, an average yaw error of $10°$ can cause a power loss up to 10%. The cosine cube law is discussed also in [5], where it is argued that the law is rather $\cos^X$ with X depending on the operation regime and ranging from 1.88 to 5.14. In [6], the aeroelastic damping under yawed conditions has been analyzed and it has been observed that the aeroelastic stability of the horizontal axis wind turbine blade is adversely affected by the effects of the yaw errors: therefore, it can be stated that the yaw error not only impacts on the energy yield, but also potentially on the wind turbine lifetime. Wind turbines in operation are analyzed in [7] and the point of view is that the energy capture characteristics must be investigated under the joint effect of non-vanishing yaw angle and internal wind turbine control: a yaw index in relation to the power curve is formulated and tested on the operation data of a 2 MW wind turbine and the main result is that the yaw effect has an increasing trend up to 8 m/s and has a decreasing trend thereon because, approaching rated power, the pitch control has a more important effect with respect to yaw control.

The results in [7] provide a clear explanation of the fact that wind turbine yaw control optimization has remarkable potentiality for wind energy technology development. It should be noticed that this issue is at least twofold: on one hand, the point can be improving the control algorithm; on the other hand, the point can be improving the measurement chain basing on which the control algorithm operates. Actually, the use of cup anemometers placed behind the rotor limits the quality of wind direction measurements and in [8] it is estimated that this can negatively affect the power production up to 5%; the necessity of upwind (instead of downwind) wind flow measurements for improving yaw alignment is discussed also in [9]. The design of optimized yaw controls has been addressed in several studies, like [10–15]. An interesting aspect is that, in recent years, it has become a common practice to update the technology of operating wind turbines and this also involves the yaw control: the study in [16] is devoted to the estimate of energy yield improvement in an industrial wind farm, where a pilot wind turbine has installed a yaw control optimization. The control upgrade considered in [16] acts by diminishing the operation time with high yaw misalignment and the estimate provided in that study is that the AEP improves up to the order of 1%. For completeness, it should be mentioned that yaw control is considered the keystone for wind farm control and wake redirection [17–20].

The above cited studies deal with the dynamic analysis of the yaw error, but in wind turbines practice it can happen that, for reason of installation or operation in a complex environment, there is a systematic zero-point shift in the yaw angle measurements. If such a systematic error occurs, the control system will regulate the wind turbine to achieve a set point that is believed to be $0°$, but it is not. Therefore, if the zero-point shift of the yaw angle reaches the order of several degrees, the energy yield of the wind turbine will be systematically affected up to the order of some percent of the AEP.

The systematic yaw error can most reliably be diagnosed if additional sensory systems for measuring upwind wind flow are employed: for example, the studies in [21–24] deal with this objective through the use of Lidar anemometers. The point with this approach is its application in wind energy practice: employing a Lidar is costly and the outcome about the presence of a relevant systematic yaw error is uncertain. For this reason, a recommended practice is employing in the most intelligent way the data that are available at zero cost, i.e., the Supervisory Control And Data Acquisition (SCADA) data [25,26]. There are some studies dealing with the use of SCADA data for systematic yaw error diagnosis: in [27], the approach is analyzing the power output of the target wind turbine as a function of the yaw error; in [28], the power factor C_p (rather than the power) is selected as indicator. In [29], a wind farm approach has been developed: it is based on comparing the distribution of yaw errors for each wind turbine against the average of the wind farm.

On these grounds, the objectives of the present study are mainly two:

- contributing to the SCADA-based methodologies for diagnosing systematic yaw error;
- assessing the impact of systematic yaw error through innovative techniques based on SCADA data analysis.

The above objectives can be achieved in the present study because a very appropriate test case study has been selected. This work has been organized as a cooperation between the University of Perugia and the Renvico company. A multi-MW wind turbine sited in southern Italy, owned by the Renvico company, has been diagnosed of systematic yaw error through the SCADA-based method proposed in this work.

The innovation proposed in this work, as regards the diagnosis of systematic yaw error through operation data, is that the rotor speed has been selected as a target because it does not depend on the nacelle anemometer measurements, it only depends on the torque exerted on the rotor and it is the fundamental quantity at the basis of wind turbine control [30]. The rationale for this choice is therefore that torque is lost if the yaw angle is systematically non-zero when it should be zero and, by analyzing the symmetry of the yaw error–rotor speed curve at the target wind turbine and by comparing the curve for the target and for reference wind turbines, it is expected that the yaw error diagnosis should be effective and clear.

Subsequently, for corroborating the SCADA-based analysis, upwind wind flow measurements have been conducted using spinner anemometer [31] and a 4° systematic yaw error has been corrected by the wind turbine manufacturer. Therefore, for the present study, operation data before and after the correction of 4° yaw error have been at disposal and this has allowed studying how the performance of the target wind turbine has changed.

A further innovation of this study deals with the estimation of energy yield improvement: it has been pursued through the approach employed, for example, in [32–34]. The rationale is that in wind farm operation there are situations when the performance of a target wind turbine is expected to change, while there is no reason why the performance of the remainder wind turbines should change: this happens for example when a technology upgrade is installed on a pilot wind turbine [32–34], or when a systematic error at a wind turbine is corrected [35]. Therefore, the change in the relative performance between the target wind turbine and the reference wind turbines can be codified by employing the operation variables of the reference wind turbines as inputs for a data-driven model whose output is the power of the target wind turbine. The correction of the yaw error at one wind turbine is expected to produce a performance improvement and therefore, if the model is trained with pre-correction data, validating the model on pre and post correction data sets, different behaviors in the residuals between measurements and model estimates should be observed.

Summarizing, therefore, with this study it is supported that the use of operation data for diagnosing systematic wind turbine yaw errors is fruitful and that this effort is strongly encouraged in the wind energy practice because the correction of 4° yaw error provided an energy yield improvement of the order of 1.5% of the AEP on a multi-MW wind turbine.

The structure of the manuscript is the following: in Section 2, the test case and the data sets at disposal are described; Section 3 is devoted to the methods for yaw error detection and for performance assessment; the results are collected and discussed in Section 4; conclusions and further directions are indicated in Section 5.

2. The Test Case and the Data Set

The wind farm features six multi-MW wind turbines, it is sited in southern Italy and the owner company is Renvico. The lowest inter-turbine distance is of the order of more than six rotor diameters.

The target wind turbine for the diagnosis of systematic yaw error is WTG05 and it is indicated in red in the layout reported in Figure 1.
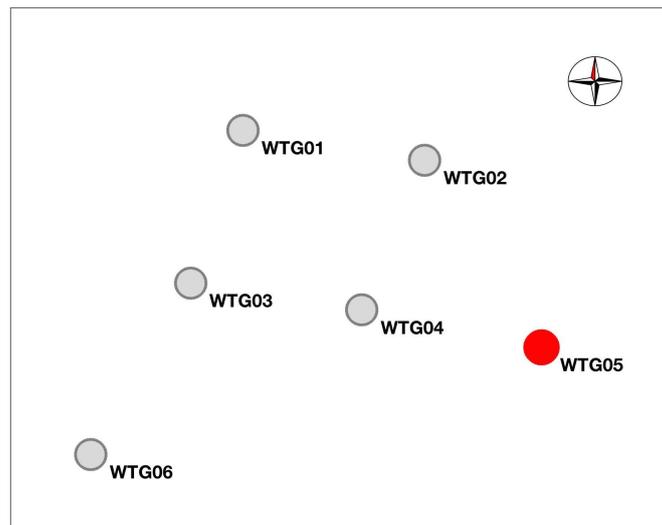
Figure 1. The layout of the wind farm.

A possible systematic yaw error on WTG05 has been at first diagnosed using the methods proposed in this study; subsequently, the results based on operation data analysis have been corroborated by upwind wind flow measurements from a spinner anemometer [31], and in the last days of 2019 the wind farm manufacturer has corrected a 4° systematic yaw error.

The data sets at disposal are the following:

- D_{bef} goes from 1st January 2017 to 1st January 2018: it is a data set prior to the yaw error correction on WTG05;
- D_{aft} goes from 1st January 2020 to 1st April 2020: it is a data set posterior to the yaw error correction on WTG05.

The validated data at disposal, for each wind turbine in the wind farm, are:

- nacelle wind speed v;
- nacelle wind direction θ_{wd};
- nacelle position θ_{nac};
- power production P;
- rotor speed ω;
- generator speed Ω;
- blade pitch angle β.

The data have 10 min of sampling time and the wind turbine grid operation time counters were available too and have been used to filter data on the request that the wind turbines of interest were operating properly and producing output. Upon data filtering, D_{bef} results consisted of 13,569 records and D_{aft} of 4915 records.

3. Methods

3.1. Yaw Error Detection

The general idea proposed in this work for the diagnosis of systematic yaw error is analyzing the yaw error–rotor speed curve. As hinted in Section 1, the rationale for the selection of this target is that the rotor speed of a wind turbine does not depend on the wind flow measurements from the nacelle anemometer placed behind the rotor span. The wind turbine regulates the rotor speed basing on the torque exerted on the rotor and the wind turbine control follows consequently. Therefore, if a wind turbine is losing torque because there is a systematic yaw error, it is losing rotor speed.

The yaw error is defined in Equation (1)

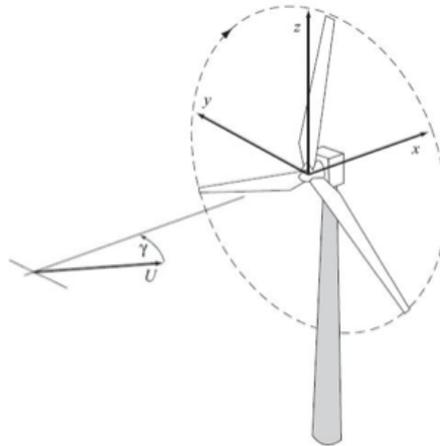$$\gamma = \theta_{nac} - \theta_{wd} \tag{1}$$

and represented in Figure 2.

Figure 2. Representation of the yaw error γ.

The proposed procedure is the following:

- filter the data on wind turbine operation time, using the appropriate time counter;
- filter the data on a narrow wind speed interval in Region II.
- compute the average yaw error–rotor speed curve. Data can be averaged on yaw error intervals of, for example, 1° or 2°.

According to the reasoning in [7], it is appropriate to filter data on a wind speed interval that is not too close to rated power because, approaching rated, the pitch control is more important than the yaw control and because the rotor speed saturates. For a multi-MW wind turbine, an appropriate choice can be for example [7, 8] m/s.

Two aspects of the yaw error–rotor curve provide meaningful indications about the presence of a systematic yaw error and these, summarizing, are the symmetry of the curve and the comparison between the curves of the wind turbines in the farm:

- the curve should have its maximum for $\gamma \simeq 0$ and diminish for increasing absolute value of γ. On the other way round, if there is a systematic yaw error, the curve is clearly asymmetric and has its maximum at the value of γ corresponding to the systematic error.
- if all the wind turbines in the farm are well aligned, the observed average ω_r should have the same order of magnitude for all the wind turbines. If a wind turbine is losing rotor speed, it is under-performing.

The above two aspects can be summarized with two indexes: an asymmetry index η_{asym} and a discrepancy index η_{disc}. The former can be defined, given the yaw error–rotor rpm curve for the target wind turbine, as the difference between the average rotor rpm for $\gamma < 0$ (ω_-) and for $\gamma > 0$ (ω_+) as in Equation (2):

$$\eta_{asym} = \bar{\omega}_- - \bar{\omega}_+. \tag{2}$$

The discrepancy index η_{disc} is defined in Equation (3):

$$\eta_{disc} = \bar{\omega}_{ref} - \bar{\omega}_{tar}, \tag{3}$$

where $\bar{\omega}_{ref}$ is the average rotor speed of one (or more) reference wind turbine and $\bar{\omega}_{tar}$ is the average rotor speed of the target wind turbine.

Notice that the above indexes can be computed on the data sets D_{bef} and D_{aft}: if the line of reasoning is correct, after the yaw error correction (therefore in D_{aft}), the indexes should highlight a remarkable difference.

3.2. Support Vector Regression and Data Set Analysis

The objective of this part of the work is formulating a reliable method for quantifying the performance improvement of wind turbine WTG05 after the $4°$ yaw error correction. This translates in the necessity of precision wind turbine performance monitoring through operation data analysis: this is in general a complex task because the power of a wind turbine has a multivariate dependence on climate conditions and working parameters.

The power curve in its simplest and widely accepted conception [36] can often provide meaningful indications about wind turbine performance, but in general it can be vexed by season bias [37], wind profile [38] and by data quality issues, due to the fact that the wind speed is measured through cup anemometers placed behind the rotors.

For this reason, there is a vast amount of literature about the modelling of wind turbine power: in [39,40], the drawbacks of the standard binned power curve modeling method [36] are discussed and it is argued that more input variables (as, for example, rotor speed) and a more sophisticated modelling procedure (as, for example, Gaussian process) are needed for precision modelling of wind turbine power.

The point behind the intuition in [32–34] is that in wind energy practice there are situations where it is known that one (or more) wind turbines are expected to change their performance, while the rest of the wind farm is not. An example is when a technology upgrade is installed on a pilot wind turbine from a wind farm and only at a later stage, upon assessment of the energy yield improvement, the decision about extending the upgrade to all the wind farm is taken. Another similar situation is when a systematic error [35] (as in the present test case) is corrected on a wind turbine: after the yaw error correction of $4°$, WTG05 is expected to improve its performance, while the other wind turbines are not.

The idea is therefore that in this case it is possible not only to improve the model structure (with respect to binned power curves), but also to enrich the model inputs on a wind farm level because some reference wind turbines can be selected, whose operation parameters can be input variables for a regression for the power of the target wind turbine of interest (WTG05, in this case). This can be conceived as a wind farm generalization of the concept of rotor equivalent wind speed [41,42].

For the present test case, WTG02, WTG04 and WTG06 have been selected as reference wind turbines because, on the grounds of the knowledge of the history of the interventions on the wind farm, it has been evaluated that they are the most appropriate.

The input variables have been selected basing on their Pearson correlation coefficient with the target Y (power of WTG05). Given in general a pair of variables x and y, the Pearson correlation coefficient is defined in Equation (4):

$$r_{xy} = \frac{N \sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{\sqrt{N \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2} \sqrt{N \sum_{i=1}^{N} y_i^2 - \left(\sum_{i=1}^{N} y_i\right)^2}},$$ (4)

where N is the number of x and y observations.

Basing on this criterion, the following input variables have been selected for the model whose target Y is the power of WTG05:

- power of WTG02: P_{WTG02};
- power of WTG04: P_{WTG04};
- power of WTG06: P_{WTG06};

- rotor speed of WTG02: ω_{WTG02};
- rotor speed of WTG04: ω_{WTG04};
- rotor speed of WTG06: ω_{WTG06};
- generator speed of WTG02: Ω_{WTG02};
- generator speed of WTG04: Ω_{WTG04};
- generator speed of WTG06: Ω_{WTG06}.

These variables constitute the covariates matrix $\boldsymbol{X}$.

In Table 1, the statistical properties of the covariates matrix and the Pearson coefficient with the output for the data set D_{bef} are reported. The data set D_{bef} is further represented in Figure 3, where the time series of the covariates and of the output are reported: these are grouped in subfigures, each representing homogeneous quantities (i.e., powers, rotor speeds, generator speeds of the reference wind turbines; power of the target wind turbine). Furthermore, in Figure 4, two scatter plots are reported for the data set D_{bef}: the power of WTG05 is scattered against the powers and the rotor speeds respectively of the reference wind turbines.
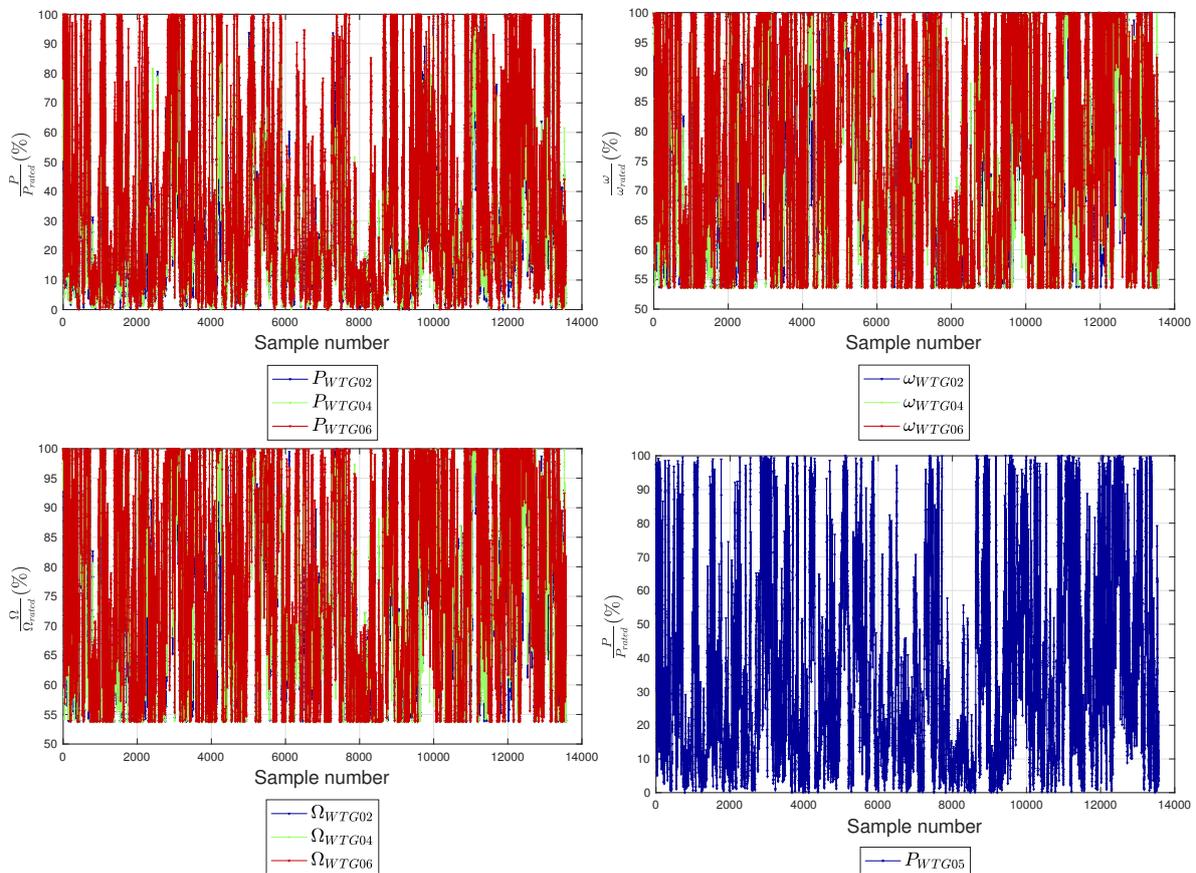

Figure 3. Time series representation of the D_{bef} data set: from the upper left to the lower right, respectively the powers, the rotor speeds, the generator speeds of the reference wind turbines, the power of the target wind turbine WTG05. All the measurements are represented in units of the rated.
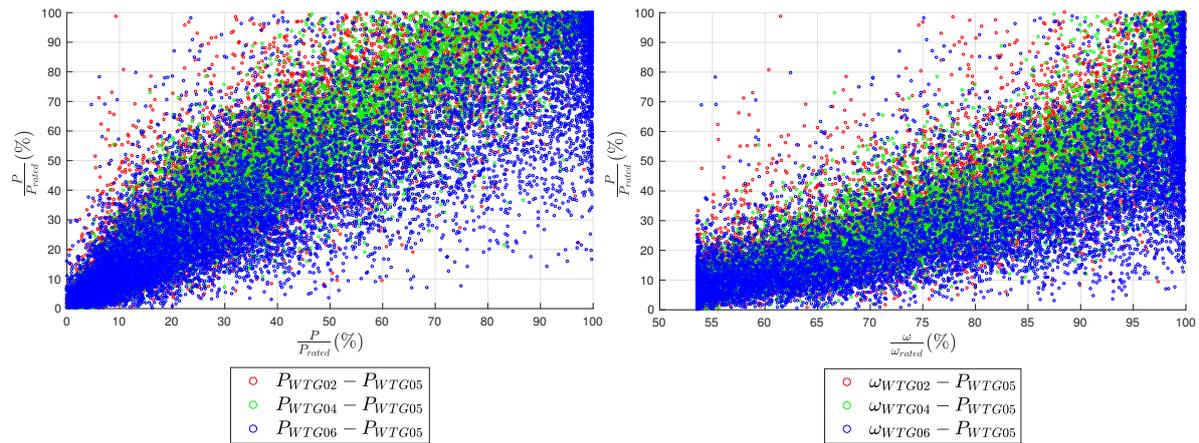
Figure 4. Scatter plot representation of the D_{bef} data set: in *y*-axis, the power of the target wind turbine WTG05 and in the *x*-axis the powers of the reference wind turbines (**left**), the rotor speeds of the reference wind turbine (**right**). All the measurements are represented in units of the rated.

Table 1. Statistical indicators of the covariates matrix X for the data D_{bef} and Pearson coefficient with the output.

Covariate	Mean (% of Rated)	Coefficient of Variation (%)	Skewness	Kurtosis	r_{xy}
P_{WTG02}	29.5	85.1	1.0	3.0	0.91
P_{WTG04}	30.9	84.2	0.9	2.9	0.94
P_{WTG06}	38.2	83.3	0.7	2.1	0.87
ω_{WTG02}	75.0	21.9	0.1	1.6	0.90
ω_{WTG04}	75.8	22.0	0.09	1.5	0.91
ω_{WTG06}	78.9	22.2	-0.1	1.5	0.82
Ω_{WTG02}	75.1	21.8	0.2	1.6	0.89
Ω_{WTG04}	75.9	21.9	0.09	1.5	0.92
Ω_{WTG06}	78.9	22.1	-0.1	1.5	0.84

In order to appreciate qualitatively the performance improvement of WTG05 after yaw error correction, in Figure 5, the power of WTG05 has been represented as a function of the power of WTG02 for the data sets D_{bef} and D_{aft}: on the left, the scatter plot is represented and on the right the average curve, with data averaged in interval of the 10% of the rated power, is reported.
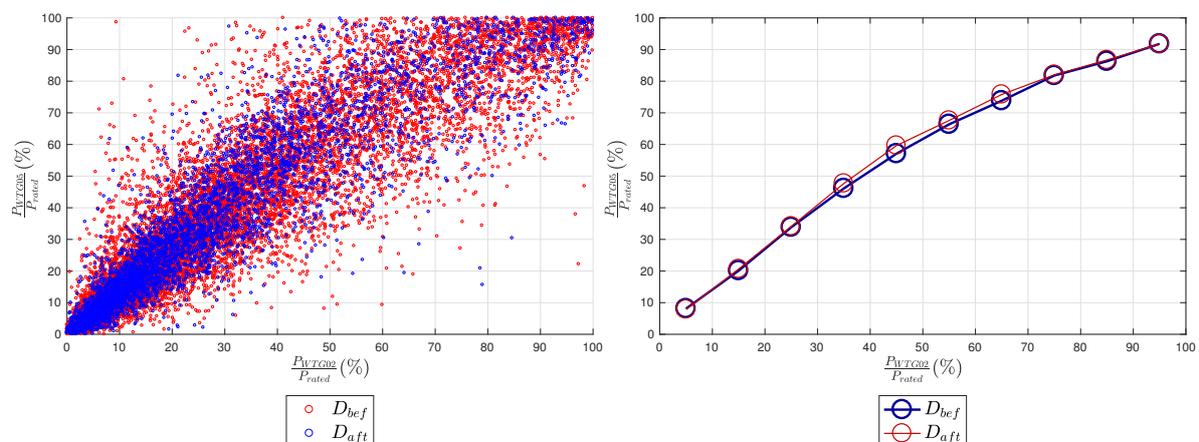

Figure 5. Representation of the P_{WTG06} as a function of P_{WTG02} for data sets D_{bef} and D_{aft}: scatter plot (**left**) and average curve (**right**).

From the plot on the right in Figure 5, it qualitatively arises that the performance of WTG05 with respect to WTG02 slightly increases during D_{aft}, in particular in Region II as expected. Furthermore, it arises that the relation between the power of WTG05 and the power of WTG02 is in general non-linear.

A coherent procedure for estimating the performance improvement is needed, taking into account the multivariate non-linear dependence of the power of WTG05 on the model inputs: for this reason, the selected model structure is Support Vector Regression with Gaussian Kernel.

In order to understand the principles of Support Vector Regression [43], consider at first a linear model (Equation (5)):

$$f(X) = X\beta + b. \tag{5}$$

The objective is finding $f(X)$ with the minimum norm value $\beta'\beta$ subject to the residuals being lower than a threshold ϵ for each observation (Equation (15)):

$$|Y_n - X_n\beta + b_n| \leq \epsilon \tag{6}$$

The optimization of the regression is a compromise between the flatness of $f(x)$ and tolerance on the occurrence of residuals higher than ϵ. This problem can be rephrased through the Lagrange dual formulation: the function to be minimized is $L(\alpha)$ (Equation (7)):

$$L(\alpha) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) X_i'X_j + \epsilon \sum_{i=1}^{N} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} Y_i(\alpha_i^* - \alpha_i), \tag{7}$$

with the constraints (Equation (8))

$$\sum_{n=1}^{N} (\alpha_n - \alpha_n^*) = 0$$

$$0 \leq \alpha_n \leq C \tag{8}$$

$$0 \leq \alpha_n^* \leq C,$$

where C is the box constraint.

The β parameters are given in Equation (9):

$$\beta = \sum_{n=1}^{N} (\alpha_n - \alpha_n^*) X_n. \tag{9}$$

If α_n or α_n^* are non-vanishing, the corresponding observation is called a support vector.

Once the model has been trained with a data set, it can be used for predicting new values, given the inputs, through the function (Equation (10)):

$$f(X) = \sum_{n=1}^{N} (\alpha_n - \alpha_n^*) X_n'X + b. \tag{10}$$

Equation (10) is fundamental for the objectives of the present work because the steps for performance improvement characterization are:

- train the model on a pre-correction data set;
- predict new values on a pre-correction data set;
- predict new values on a post-correction data set.

The production improvement can be elaborated by studying how the discrepancy between predicted new values and observed new values changes after the yaw error correction.

For this work a non-linear regression has been selected: in general, it is obtained by replacing in the above formulas the dot products between observations matrix with a nonlinear Kernel function (Equation (11)):

$$G(X_1, X_2) = \langle \varphi(X_1)\, \varphi(X_2) \rangle, \tag{11}$$

where φ is a transformation mapping the X observations into a feature space.

In this work, a Gaussian Kernel (Equation (12)) has been selected:

$$G\left(X_i, X_j\right) = e^{-\left\|X_i - X_j\right\|^2}. \tag{12}$$

Then, for the nonlinear case, Equation (7) rewrites as in Equation (13):

$$L\left(\alpha\right) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right) G\left(X_i, X_j\right) + \epsilon \sum_{i=1}^{N} \left(\alpha_i + \alpha_i^*\right) + \sum_{i=1}^{N} Y_i\left(\alpha_i^* - \alpha_i\right), \tag{13}$$

and Equation (10) for predicting rewrites as in Equation (14):

$$f(X) = \sum_{n=1}^{N} \left(\alpha_n - \alpha_n^*\right) G\left(X_n, X\right) + b. \tag{14}$$

3.3. Performance Analysis

The data sets at disposal are employed as follows for estimating the energy yield improvement subsequent to the yaw error correction:

- D_{bef} is randomly divided into two subsets: D0 (a random selection of $\frac{2}{3}$ of the data set) and D1 (the remainder $\frac{1}{3}$ of the data set). D0 is used for training the regression, D1 is used for testing the regression. The convergence of model training is verified through the MATLAB® fitrsvm routine.
- D_{aft} (also named D2 for notation consistency) is used to quantify the performance deviation with respect to D1 (and therefore D_{bef}).

Once the Support Vector Machine model has been trained with the D0 data set, the output is simulated using Equation (14) (basing on the input variables observations) for the data sets D1 and D2. If the performance of the wind turbine has improved in D2 with respect to D1, the behavior of the difference between measured and simulated target should have changed

Therefore, consider Equation (15) with $i = 1, 2$.

$$R(X_i) = Y(X_i) - f(X_i). \tag{15}$$

For $i = 1, 2$, one computes (Equation (16))

$$\Delta_i = 100 \frac{\sum_{X \in Data_i} \left(Y(X) - f(X)\right)}{\sum_{X \in Data_i} Y(X)} \tag{16}$$

and the quantity $\Delta = \Delta_2 - \Delta_1$ provides an estimate of the performance deviation from data set D1 to D2 [44,45]. In other words, if during D1 and D2 the wind turbine performance does not change significantly, the residuals (Equation (15)) have similar statistical properties: therefore, $f(X)$ in average will deviate from $Y(X)$ in a similar way in D1 and D2 and therefore Δ_2 and Δ_1 will be similar and their difference will approximately be 0. If, instead the performance improves during D2, the residuals (Equation (15)) are on average higher than in D1, because $f(X)$ is estimated through a model trained on a data set characterized by a lower performance. Therefore, one would reasonably obtain Δ non-negligibly higher than 0.

4. Results

4.1. Yaw Error Diagnosis

The method described in Section 3.1 has been applied to the data sets D_{bef} and D_{aft}. Results are reported for the [7, 8] m/s wind speed interval selection, but it has been crosschecked that the method does not depend on this peculiar selection, as long as rated power (and therefore rotor speed saturation) is not reached.

In Figures 6 and 7, WTG05 has been compared to a sample reference wind turbine (WTG06) in respectively the data set D_{bef} and D_{aft}.
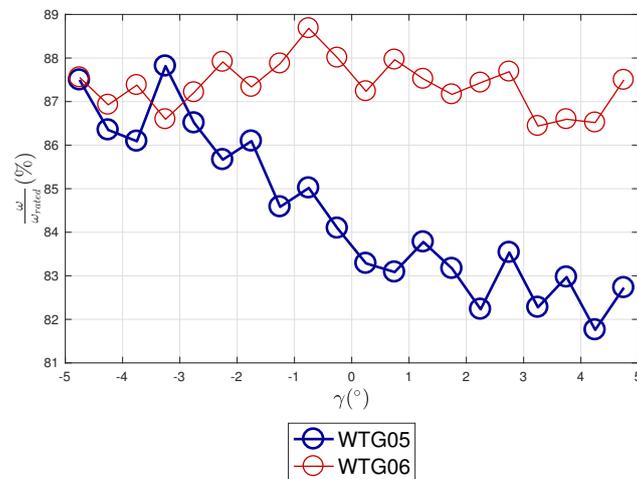

Figure 6. The yaw error–rotor speed curve for WTG05 and a sample reference wind turbine (WTG06): [7, 8] m/s wind speed interval. Data set D_{bef}.
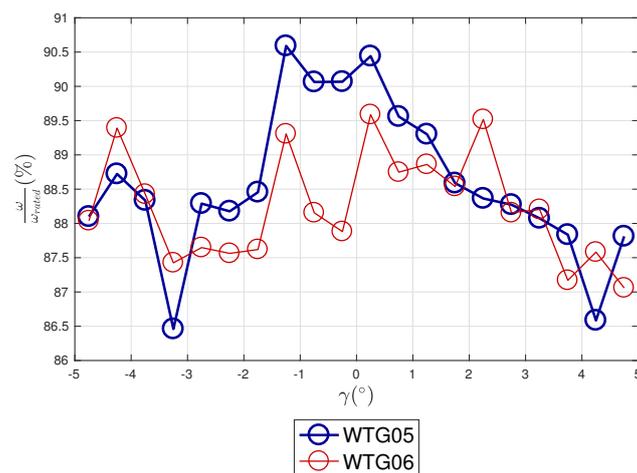

Figure 7. The yaw error–rotor speed curve for WTG05 and a sample reference wind turbine (WTG06): [7, 8] m/s wind speed interval. Data set D_{aft}.

Comparing Figure 6 and Figure 7, it is possible to observe that Figure 6 describes the wind turbine WTG05 affected by a systematic yaw error. This can be argued because in Figure 6, the highest rotor speed is achieved when the yaw error is of the order of 5° and the curve is clearly asymmetric. Furthermore, the highest torque should be exerted on the wind turbine when the yaw error is around 0° and it is therefore expected that the curves of WTG05 and WTG06 should be comparable, especially at 0° yaw error when there is the highest data population: instead, it happens that around 0°, the rotor speed deficit for WTG05 with respect to WTG06 is of the order of the 10% of the rotor speed working range for that model of wind turbine.

This kind of analysis has inspired further investigation of the present test case and a spinner anemometer has been installed on site for upwind wind flow measurements. The spinner anemometer data analysis has corroborated the present conclusions and a systematic yaw error of 4° has been corrected by the wind turbine manufacturer.

D_{aft} therefore describes WTG05 after the yaw error correction and the expectation is that the yaw error–rotor speed curves for WTG05 and the reference wind turbine (WTG06) should be comparable.

This is indeed the case, as can be seen in Figure 7: the first indication is that the working rotor speed ranges of WTG05 and WTG06 are very similar and the the WTG05 curve is not severely asymmetric as it was in Figure 6. In other words, from Figure 7 the observation is that WTG05 is not severely losing rotor speed (and therefore torque) as before the yaw error correction.

The above considerations can be summarized in Table 2, displaying how the asymmetry index η_{asym} (Equation (2)) and the discrepancy index η_{disc} (Equation (3)) change from D_{bef} to D_{aft}. It results that both indexes are of the order of 10 times higher in D_{bef} with respect to D_{aft}.

Table 2. Asymmetry index η_{asym} (Equation (2)) and Discrepancy index η_{disc} (Equation (3)) before and after yaw error correction.

Data Set	η_{asym}	η_{disc}
D_{bef}	3.3	2.9
D_{aft}	0.3	−0.3

As far as can be argued through operation data analysis, therefore, it can be stated that WTG05 has recovered the normal operation after the yaw error correction.

4.2. Performance Assessment

As a preliminary analysis, the SVM regression has been subjected to *K*-fold cross-validation [46]. $(K-1)/K$ of the data are used for training and the remaining $1/K$ are used for validation. *K* has been selected to be 10, as typical in this kind of applications. The models store the results of training on nine folds of observations out of 10, and leave one fold of observations out for validation. The generalization error can be estimated as the out-of-sample root mean-squared error and the result is 62 kW. Subsequently, all the training data set D0 has been passed for the training and the convergence of the model has been achieved after 5761 iterations. The resubstitution (in-sample) root mean-squared error is 53 kW.

Following Equation (15), for $i = 1, 2$, one can write the average residual between measurements and model estimates as

$$\bar{R}_i = \frac{1}{N_i} \sum_{X \in Di} Y(x) - f(X), \qquad (17)$$

where N_i is the sample size in D1 and D2.

The WTG05 performance improvement after the yaw error correction should resemble in different statistical properties of the residuals $R(X_2)$ with respect to $R(X_1)$. If the model $f(X)$ is trained with pre-correction data, when the output is simulated on the post-correction data set D2, the model estimate $f(X)$ should be systematically lower than the measurements in a manner which should be distinguishable with respect to D1. This is indeed the case, as can be seen in Table 3: the measured WTG05 power is averagely 13.4 kW higher than the model estimate after the yaw correction with respect than before the correction. It should be noticed that the selected data-driven approach allows synthesising the experiments replication by running the model with different random choices of D0 (and consequently of D1). The results in Table 3 and all the following results should therefore be intended as the average results of 1000 iterations, but it should be noticed that few iterations are sufficient for obtaining a substantially stable results.

From this analysis, the obtained result is $\Delta = 1.9\%$. Recalling that data have been filtered below rated power, because no performance improvement can be seen when the power is already rated, the result for Δ means that on average WTG05 has improved its performance of the 1.9% below rated power after the correction of the yaw error. This result should converted in an estimate for the AEP improvement; for a detailed analysis about this aspect in relation to the Weibull wind intensity distribution, refer to [2]. For the purposes of the present work, a simple method can be employed: it is based on the analysis of the wind intensity statistics on site, because the resulting improvement in

the AEP depends on how much time the wind turbine operates at rated power in that particular site. Therefore, Equation (18) provides a simple yaw to estimate Δ_{AEP}:

$$\Delta_{AEP} = \frac{\Delta \cdot E_{sub-rated}}{E_{sub-rated} + E_{rated}}, \tag{18}$$

where $E_{sub-rated}$ is the yearly sub-rated energy yield and E_{rated} is the yearly rated energy yield. Equation (18) has been employed on several yearly test data sets and it results that the order of magnitude for Δ_{AEP} is 1.5%.

Table 3. Statistical behavior of the residuals between measurement and estimation, for the D0 and D1 data sets.

Data Set	$\bar{R}$ (kW)
D1	−0.9
D2	12.5

Finally, it is interesting to analyze the dependence of the energy yield improvement for the different operation regions of the wind turbines. This can be appreciated in Figure 8: $R(X_1)$ and $R(X_2)$, computed on a sample model run, are displayed. The data are averaged in power production intervals, whose amplitude is 10% of the rated. From Figure 8 it arises that the correction of the yaw error improves the wind turbine performance especially between 35% and 75% of the rated power, with an increasing effect with increasing power. This corroborates the observations in [7] that the yaw behavior of a wind turbine must be analyzed taking into account the control and the operation regime. When rated power is approached, the pitch behavior becomes important and the yaw error becomes less relevant. Basing on the results in Figure 8, it is also supported that the yaw error is less relevant near the cut-in, because in this regime the control also involves the pitch. The yaw error is therefore particularly relevant in Region II, when the rotor speed increases and the blade pitch is practically constant. From the estimate about Δ_{AEP} and the result in Figure 8, it is argued that, despite the effects of systematic yaw error are concentrated in Region II, they are definitely non-negligible with respect to the energy yield of the wind turbine. Furthermore, it is a useful crosscheck to observe that the quantitative result of Figure 8 is coherent with the qualitative analysis of Figure 5.
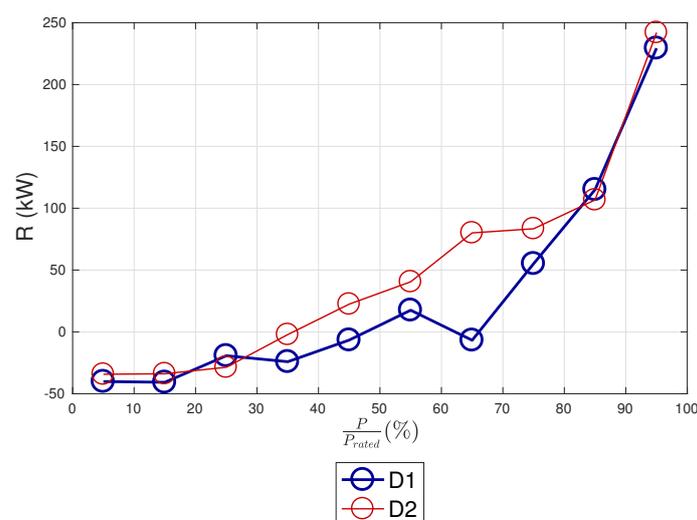

Figure 8. The average difference R between power measurement Y and estimation $f(X)$ (Equation (15)). Data sets: D1 and D2. Sample run of the model.

5. Conclusions

The present study has been devoted to innovative operation data analysis methodologies for the diagnosis of wind turbine systematic yaw error and for the assessment of the impact of this kind of error on wind turbine energy yield. It has been organized as a collaboration between academia (University of Perugia) and industry (Renvico company) and it has the structure of a test case discussion.

The selected test case is a multi-MW wind turbine sited in Italy that has been diagnosed of a remarkable systematic yaw error through the operation data analysis of the present work. These results have been corroborated by upwind flow measurements through spinner anemometer and a systematic yaw error of 4° has been corrected. Therefore, operation data describing the wind turbine in operation before and after the correction of this yaw error have been at disposal for this study and this has allowed estimating how much the wind turbine performance has improved after the yaw error correction or, on the other hand, how much energy yield is lost when there is a yaw error.

The two main results of this study are the following:

- Despite the fact that SCADA data do not provide upwind flow measurements, it is nevertheless possible (and recommended) to employ them for reliably diagnosing systematic yaw errors. The main innovation of this study has been targeting the rotor speed, because in Region II it directly depends on the torque exerted on the rotor and torque is lost if the yaw angle is systematically non-zero when it should be zero. Analyzing the properties of the yaw error–rotor speed curve, for each single wind turbine and comparing wind turbines in a wind farm, it is possible to obtain meaningful indications about the presence of systematic yaw errors.
- As discussed in [7], the effect of yaw error should be read in light of the wind turbine control. The energy yield is affected by the systematic yaw error in particular in Region II, when the torque and the rotor speed increase with the wind and the blade pitch angle is practically set to 0. Near the cut-in and near rated power, the importance of pitch control increases and the yaw error is less important. The net effect of this on wind turbine operation can be estimated from the test case considered in this work: a systematic yaw error of 4° can affect the AEP of the wind turbine for the order of 1.5%.

Furthermore, an interesting aspect of the present work is that it contributes in general to the problem of wind turbine performance monitoring through operation data analysis. It is well known in the literature that the power curve binning method (as recommended by the IEC [36]) is simplistic for precision monitoring and that reliable data-driven models for the power of a wind turbine should have a more complicated structure (as, for example, PCR, GP, LASSO regression and so on) and most of all should be multivariate, in order to account for the complex dependence of wind turbine power on climate conditions and working parameters.

The present study has contributed to the above issue with the innovative intuition that, in the wind farm practice, there are situations in which it is known that the performance of some wind turbines could (or should) have changed and that there is no reason why the performance of other wind turbines should have changed. The idea of this study is that this information should be exploited for the data-driven modelling of the target wind turbine of interest and the input variables for the model can include the operation variables of the reference wind turbines.

An interesting further direction of the present work is applying the present methodologies on a vast number of wind turbines, possibly operating in field for a considerable number of years; actually, this kind of study could contribute in shedding light on the occurrence of systematic yaw errors in a wind turbine lifetime, possibly depending on installation, human management interventions, wind turbine technology, and operation in a harsh environment.

Author Contributions: Conceptualization, D.A, F.C., M.B., A.L. and L.T. ; data curation, D.A. and A.L.; formal analysis, D.A. and F.C.; investigation, D.A.; methodology, D.A.; project administration, F.C. and L.T.; software, D.A.; supervision, F.C., A.L. and L.T., validation, D.A., F.C., M.B., A.L. and L.T.; writing—original draft, D.A.;

Nomenclature

List of the mathematical symbols.

Symbol	Meaning
v	Nacelle wind speed
θ_{wd}	Nacelle wind direction
θ_{nac}	Nacelle position
P	Power production
ω	Rotor speed
Ω	Generator speed
β	Blade pitch angle
γ	Yaw error
η_{asym}	Asymmetry Index
η_{disc}	Discrepancy Index
r	Pearson coefficient
R	Residuals between measurement and model estimate
Δ	Average percentage residual
E	Energy Yield

List of the abbreviations.

Acronym	Meaning
AEP	Annual Energy Production
GP	Gaussian Process
HAWT	Horizontal-Axis Wind Turbine
IEC	International Electrotechnical Commission
LASSO	Least Absolute Shrinkage and Selection Operator
PCR	Principal Component Regression
SCADA	Supervisory Control And Data Acquisition
SVM	Support Vector Machine
SVR	Support Vector Regression
WTG	Wind Turbine Generator

References

1. McKenna, R.; vd Leye, P.O.; Fichtner, W. Key challenges and prospects for large wind turbines. *Renew. Sustain. Energy Rev.* **2016**, *53*, 1212–1221. [CrossRef]
2. Saenz-Aguirre, A.; Fernandez-Gamiz, U.; Zulueta, E.; Aramendia, I.; Teso-Fz-Betono, D. Flow control based 5 MW wind turbine enhanced energy production for hydrogen generation cost reduction. *Int. J. Hydrog. Energy* **2020**. [CrossRef]
3. Burton, T.; Jenkins, N.; Sharpe, D.; Bossanyi, E. *Wind Energy Handbook*; John Wiley & Sons: Hoboken, NJ, USA, 2011.
4. Wan, S.; Cheng, L.; Sheng, X. Effects of yaw error on wind turbine running characteristics based on the equivalent wind speed model. *Energies* **2015**, *8*, 6286–6301. [CrossRef]
5. Schulz, C.; Letzgus, P.; Lutz, T.; Krämer, E. CFD study on the impact of yawed inflow on loads, power and near wake of a generic wind turbine. *Wind Energy* **2017**, *20*, 253–268. [CrossRef]
6. Jeong, M.S.; Kim, S.W.; Lee, I.; Yoo, S.J.; Park, K. The impact of yaw error on aeroelastic characteristics of a horizontal axis wind turbine blade. *Renew. Energy* **2013**, *60*, 256–268. [CrossRef]
7. Dai, J.; Yang, X.; Hu, W.; Wen, L.; Tan, Y. Effect investigation of yaw on wind turbine performance based on SCADA data. *Energy* **2018**, *149*, 684–696. [CrossRef]

8. Cortina, G.; Sharma, V.; Calaf, M. Investigation of the incoming wind vector for improved wind turbine yaw-adjustment under different atmospheric and wind farm conditions. *Renew. Energy* **2017**, *101*, 376–386. [CrossRef]

9. Kragh, K.A.; Hansen, M.H. Potential of power gain with improved yaw alignment. *Wind Energy* **2015**, *18*, 979–989. [CrossRef]

10. Song, D.; Fan, X.; Yang, J.; Liu, A.; Chen, S.; Joo, Y.H. Power extraction efficiency optimization of horizontal-axis wind turbines through optimizing control parameters of yaw control systems using an intelligent method. *Appl. Energy* **2018**, *224*, 267–279. [CrossRef]

11. Karakasis, N.; Mesemanolis, A.; Nalmpantis, T.; Mademlis, C. Active yaw control in a horizontal axis wind system without requiring wind direction measurement. *IET Renew. Power Gener.* **2016**, *10*, 1441–1449. [CrossRef]

12. Shariatpanah, H.; Fadaeinedjad, R.; Rashidinejad, M. A new model for PMSG-based wind turbine with yaw control. *IEEE Trans. Energy Convers.* **2013**, *28*, 929–937. [CrossRef]

13. Song, D.; Yang, J.; Fan, X.; Liu, Y.; Liu, A.; Chen, G.; Joo, Y.H. Maximum power extraction for wind turbines through a novel yaw control solution using predicted wind directions. *Energy Convers. Manag.* **2018**, *157*, 587–599. [CrossRef]

14. Saenz-Aguirre, A.; Zulueta, E.; Fernandez-Gamiz, U.; Lozano, J.; Lopez-Guede, J.M. Artificial Neural Network Based Reinforcement Learning for Wind Turbine Yaw Control. *Energies* **2019**, *12*, 436. [CrossRef]

15. Saenz-Aguirre, A.; Zulueta, E.; Fernandez-Gamiz, U.; Ulazia, A.; Teso-Fz-Betono, D. Performance enhancement of the artificial neural network–based reinforcement learning for wind turbine yaw control. *Wind Energy* **2019**.

16. Astolfi, D.; Castellani, F.; Natili, F. Wind Turbine Yaw Control Optimization and Its Impact on Performance. *Machines* **2019**, *7*, 41. [CrossRef]

17. Campagnolo, F.; Petrović, V.; Bottasso, C.L.; Croce, A. Wind tunnel testing of wake control strategies. In Proceedings of the IEEE American Control Conference (ACC), Boston, MA, USA, 6–8 July 2016; pp. 513–518.

18. Gebraad, P.; Teeuwisse, F.; Van Wingerden, J.; Fleming, P.A.; Ruben, S.; Marden, J.; Pao, L. Wind plant power optimization through yaw control using a parametric model for wake effects—A CFD simulation study. *Wind Energy* **2016**, *19*, 95–114. [CrossRef]

19. Fleming, P.; Churchfield, M.; Scholbrock, A.; Clifton, A.; Schreck, S.; Johnson, K.; Wright, A.; Gebraad, P.; Annoni, J.; Naughton, B.; et al. Detailed field test of yaw-based wake steering. *J. Phys. Conf. Ser.* **2016**, *753*, 052003. [CrossRef]

20. Dai, J.; Yang, X.; Yang, W.; Gao, G.; Li, M. Further Study on the Effects of Wind Turbine Yaw Operation for Aiding Active Wake Management. *Appl. Sci.* **2020**, *10*, 1978. [CrossRef]

21. Kragh, K.; Hansen, M.; Mikkelsen, T. Improving yaw alignment using spinner based LIDAR. In Proceedings of the 49th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition, Orlando, FL, USA, 4–7 January 2011; p. 264.

22. Mikkelsen, T.; Angelou, N.; Hansen, K.; Sjöholm, M.; Harris, M.; Slinger, C.; Hadley, P.; Scullion, R.; Ellis, G.; Vives, G. A spinner-integrated wind lidar for enhanced wind turbine control. *Wind Energy* **2013**, *16*, 625–643. [CrossRef]

23. Zhang, L.; Yang, Q. A Method for Yaw Error Alignment of Wind Turbine Based on LiDAR. *IEEE Access* **2020**, *8*, 25052–25059. [CrossRef]

24. Bakhshi, R.; Sandborn, P. Maximizing the returns of LIDAR systems in wind farms for yaw error correction applications. *Wind Energy* **2020**. [CrossRef]

25. Muljadi, E.; Butterfield, C.P. Pitch-controlled variable-speed wind turbine generation. *IEEE Trans. Ind. Appl.* **2001**, *37*, 240–246. [CrossRef]

26. Ding, Y. *Data Science for Wind Energy*; CRC Press: Boca Raton, FL, USA, 2019.

27. Pei, Y.; Qian, Z.; Jing, B.; Kang, D.; Zhang, L. Data-Driven Method for Wind Turbine Yaw Angle Sensor Zero-Point Shifting Fault Detection. *Energies* **2018**, *11*, 553. [CrossRef]

28. Astolfi, D.; Castellani, F.; Terzi, L. An Operation Data-Based Method for the Diagnosis of Zero-Point Shift of Wind Turbines Yaw Angle. *J. Sol. Energy Eng.* **2020**, *142*. [CrossRef]

29. Astolfi, D.; Castellani, F.; Scappaticci, L.; Terzi, L. Diagnosis of wind turbine misalignment through SCADA data. *Diagnostyka* **2017**, *18*, 17–24.

30. Liao, M.; Dong, L.; Jin, L.; Wang, S. Study on rotational speed feedback torque control for wind turbine generator system. In Proceedings of the IEEE International Conference on Energy and Environment Technology, Guilin, China, 16–18 October 2009; Volume 1, pp. 853–856.

31. Demurtas, G.; Pedersen, T.F.; Zahle, F. Calibration of a spinner anemometer for wind speed measurements. *Wind Energy* **2016**, *19*, 2003–2021. [CrossRef]

32. Astolfi, D.; Castellani, F.; Terzi, L. Wind Turbine Power Curve Upgrades. *Energies* **2018**, *11*, 1300. [CrossRef]

33. Astolfi, D.; Castellani, F.; Fravolini, M.L.; Cascianelli, S.; Terzi, L. Precision Computation of Wind Turbine Power Upgrades: An Aerodynamic and Control Optimization Test Case. *J. Energy Resour. Technol.* **2019**, *141*, 051205. [CrossRef]

34. Astolfi, D.; Castellani, F. Wind turbine power curve upgrades: Part II. *Energies* **2019**, *12*, 1503. [CrossRef]

35. Astolfi, D. A Study of the Impact of Pitch Misalignment on Wind Turbine Performance. *Machines* **2019**, *7*, 8. [CrossRef]

36. IEC. *Power Performance Measurements of Electricity Producing Wind Turbines*; Technical Report 61400-12; International Electrotechnical Commission: Geneva, Switzerland, 2005.

37. Van den Berg, G. Wind turbine power and sound in relation to atmospheric stability. *Wind. Energy* **2008**, *11*, 151–169. [CrossRef]

38. Wagner, R.; Antoniou, I.; Pedersen, S.M.; Courtney, M.S.; Jørgensen, H.E. The influence of the wind speed profile on wind turbine performance measurements. *Wind Energy* **2009**, *12*, 348–362. [CrossRef]

39. Guo, P.; Infield, D. Wind turbine power curve modeling and monitoring with Gaussian Process and SPRT. *IEEE Trans. Sustain. Energy* **2018**, *11*, 107–115. [CrossRef]

40. Pandit, R.K.; Infield, D. Comparative assessments of binned and support vector regression-based blade pitch curve of a wind turbine for the purpose of condition monitoring. *Int. J. Energy Environ. Eng.* **2019**, *10*, 181–188. [CrossRef]

41. Wagner, R.; Ca nadillas, B.; Clifton, A.; Feeney, S.; Nygaard, N.; Poodt, M.; St Martin, C.; Tüxen, E.; Wagenaar, J. Rotor equivalent wind speed for power curve measurement–comparative exercise for IEA Wind Annex 32. *J. Phys. Conf. Ser.* **2014**, *524*, 012108. [CrossRef]

42. Scheurich, F.; Enevoldsen, P.B.; Paulsen, H.N.; Dickow, K.K.; Fiedel, M.; Loeven, A.; Antoniou, I. Improving the accuracy of wind turbine power curve validation by the rotor equivalent wind speed concept. *J. Phys. Conf. Ser.* **2016**, *753*, 072029. [CrossRef]

43. Smola, A.J.; Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **2004**, *14*, 199–222. [CrossRef]

44. Lee, G.; Ding, Y.; Xie, L.; Genton, M.G. A kernel plus method for quantifying wind turbine performance upgrades. *Wind Energy* **2015**, *18*, 1207–1219. [CrossRef]

45. Hwangbo, H.; Ding, Y.; Eisele, O.; Weinzierl, G.; Lang, U.; Pechlivanoglou, G. Quantifying the effect of vortex generator installation on wind power production: An academia-industry case study. *Renew. Energy* **2017**, *113*, 1589–1597. [CrossRef]

46. Refaeilzadeh, P.; Tang, L.; Liu, H. Cross-validation. In *Encyclopedia of Database Systems*; Springer: Berlin, Germany, 2009; pp. 532–538.