A SCADA data mining method for precision assessment of performance enhancement from aerodynamic optimization of wind turbine blades

Davide Astolfi¹, Francesco Castellani¹, Ludovico Terzi²

¹ University of Perugia - Department of Engineering, Via G. Duranti 93 - 06125 Perugia, Italy
² Renvico s.r.l., Via San Gregorio 34, 20124 Milano, Italy. www.renvicoenergy.com
E-mail: davide.astolfi@unipg.it

Abstract. The target of improving efficiency of wind kinetic energy extraction has stimulated a certain attention to wind turbine retrofitting. This kind of interventions has material and labor costs and producible energy is lost during installation. Further, the estimation of the energy enhancement is commonly provided under the hypothesis of ideal conditions that can be very different from real ones. Therefore, a precise estimation of performance improvement is fundamental. In this work, a SCADA-based method is formulated for estimating the improvement in energy production of multi-megawatt wind turbines, sited in Italy in a very complex terrain. The blades of one wind turbine in the farm have been optimized by installing vortex generators and passive flow control devices. An Artificial Neural Network (ANN) model is employed: the output is the power of the retrofitted wind turbine and the inputs are the powers of some reference nearby wind turbines. The production increase is estimated by observing how the difference between simulated and measured power output changes after the installation of the aerodynamic upgrade. The average improvement is estimated as the 3.9% of the total energy produced below rated power.

1. Introduction
Wind turbine technology and condition monitoring techniques [1] have been continuously evolving and the operational unavailability of a wind turbine is estimated nowadays to be of the order of 3% of its lifetime [2]. The target of 100% of technical availability is therefore becoming realistic and this motivates the research about further optimization of wind kinetic energy conversion efficiency. Several kinds of wind turbine retrofitting are being used and they basically divide in two groups: control system upgrades and aerodynamic retrofitting. Examples of control system upgrades are pitch angle optimization [3] and high wind speed power curve optimization through the raising of the cut-out and high wind speed cut-in velocities [4, 5]. An example of aerodynamic optimization is the installation of vortex generators on wind turbine blades [3, 6]. This kind of interventions has material and labor costs and producible energy is lost during installation. Further, the estimate of the guaranteed energy improvement is commonly provided under the hypothesis of ideal operation conditions of the wind turbines, as for example absence of wakes between nearby turbines. Real wind turbines operation conditions are commonly very different from ideal ones: operation under wakes is a standard [7, 8] and, due to the increasing diffusion of onshore wind energy, also operation in complex terrain environment...
[9] is common. Therefore, a realistic estimation of performance improvement is fundamental, in order to know if the retrofitting has an advantageous return of investment. This kind of studies is quite novel in the scientific literature. An interesting example is [6]: an academia-industry joint study is presented and the effect of vortex generator installation on wind power production of two onshore farms is quantified. On one hand, SCADA data with 10 minutes of sampling time are employed; on the other hand, high-frequency power data are employed: the estimates are shown to be similar. The present work adopts a similar philosophy, because it is the collaboration between industry (Renvico srl, owning and managing 334 MW of operating wind turbines in Italy and France) and the University of Perugia. The test case is a wind farm, sited in Italy in a very complex terrain [10, 11, 12]. The blades of one wind turbine in the farm have been optimized by the manufacturer through the installation of vortex generators and passive flow control devices. The outcome is an increase of the lift and therefore of the energy production. One key point of the study is that the retrofitted wind turbine acts as a pilot test and the motivations of the joint academia-industry study lie on the fact that the decision of adopting the retrofitting on the whole wind farm depends on a precise and realistic quantification of the benefits. For accomplishing this objective, two elements are needed: SCADA data before and after the upgrade and a consistent model for simulating how much the wind turbine would have produced without the retrofitting. In [3], for example, a kernel plus method for modeling the power output of wind turbines is adopted and employed for real test cases of power upgrade assessment (pitch angle optimization and vortex generator installation). The formulation of the model for the test case of this work is dictated by a challenging peculiarity: in the period immediately after the upgrade, the nacelle transfer function of the wind turbine hasn’t been updated. Therefore, the nacelle wind speed measurements are not reliable: a power curve analysis (according, for example, to the International Electrotechnical Commission guidelines [13]) can’t be performed and the nacelle wind speed measurements can’t be used as input to a model. Moreover, the impressive complexity of the terrain makes it prohibitive to use the nearby wind turbines or the met-mast as wind conditions reference. Nevertheless, adopting a judicious Artificial Neural Network (ANN) model, it has been possible to estimate the improvement in energy production also in this complex test case and the results are quite consistent with the estimates provided by the wind turbine manufacturer under the hypothesis of ideal conditions. Summarizing, the structure of the paper is therefore the following: first, the test case and the data set are briefly described in Section 2; subsequently the model is formulated in Section 3; the results are summarized in Section 4; conclusions and further directions are sketched in Section 5.

2. The wind farm and the data set

The layout of the wind farm is reported in Figure 1. From the contour lines, it is possible to appreciate the impressive complexity of the terrain. T7 (in red in Figure 1) is the wind turbine that has been upgraded with vortex generators and passive flow control devices. The wind turbines are multi-megawatt and the hub height is 80 meters above ground level.

1 www.renvicoenergy.com
The data sets at disposal are the following:

- \( D_{bef} \) goes from 01/01/2016 to 01/07/2017. It is a period during which the standard blade configuration was adopted.
- \( D_{aft} \) goes from 01/09/2017 to 01/01/2018. It is a period during which T7 has been operating with the improved blade configuration.

3. The method

The production increase is estimated by observing how the difference between simulated and measured power output of turbine T7 behaves before and after the installation of the aerodynamic upgrade. In order to do this, a model must be formulated. The output \( y \) of the model is the power production of T7.

As stated in Section 1, the nacelle transfer function of turbine T7 hasn’t been updated by the manufacturer after the installation of the flow control devices. Therefore, the nacelle wind speed of T7 can’t be used as input for the model because the measurements during the \( D_{aft} \) data set are not reliable. In order to appreciate this, in Figure 2, the power coefficient \( C_p = \frac{P}{\frac{1}{2} \rho A v^3} \) as computed from the SCADA data during the \( D_{aft} \) data set is reported as a function of the wind speed. \( P \) is the measured power output, \( \rho \) is the air density on site, \( A \) is the blade swept area and \( v_\infty \) is the undisturbed wind speed as reconstructed from the nacelle wind speed through the nacelle transfer function. In Figure 2, the power coefficient estimates from the data of T7 and from the data of a sample wind turbine from the rest of the wind farm (T1) are reported and it clearly arises that the measurements for T7 are implausible, because well above the Betz limit. Therefore, it must be argued that the wind speed measurements at T7 are implausible.
Figure 2. The power coefficient $C_p$, as computed from the SCADA data, vs. nacelle wind speed: T7 and a sample wind turbine (T1), $D_{aft}$ data set.

On the grounds of the above discussion, it is then necessary to adopt the wind turbines nearby T7 as references for constructing the model. Due to the complexity of the terrain, it has been considered more solid to use the power output of the nearby turbines, rather than the nacelle wind speed.

The inputs $x$ to the model are therefore the power outputs of a certain number of turbines nearby T7. The selected model is a feedforward ANN model and the selected inputs are the power outputs of wind turbines T1, T2, T3, T4, T5, T6, T8, T9. The data sets at disposal are employed as follows:

- $D_{be.f}$ is randomly divided in two subsets: D0 (1 year of data) and D1 (6 months of data). D0 is used for training the model, D1 is used for validating it.
- $D_{be.f}$ (also named D2 for simplifying the notation in the following) is used to quantify the performance improvement.

The number of neurons in the hidden layer and the number of inputs have been selected by minimizing the mean absolute error between simulation and measurements in the validation period D1. 25 neurons are thus selected.

The data sets are filtered on the condition of power output production from the T1-T9 wind turbines. Further, data are filtered on power production of T7 below the rated, because at rated power the upgrade of course has no effect. Notice that measurements corresponding to wind turbines operating when the wind blows in waked sectors [13] are not filtered away, because the objective of this work is estimating how much averagely T7 improves its power production under real working conditions, that possibly include T7 or whatever nearby wind turbine operating under wake.

4. Results
First, some results can be reported supporting the choice of the model described in Section 3. The question is: since the nacelle wind speed of turbine T7 can’t be used in the model, would it be possible to use the wind speed measured at a nearby turbine (T6, namely)? Would the uncertainties increase or decrease? To answer this question, the performances of the model of Section 3 in predicting the power of T7 in the test data set D1 have been compared with the performances of an ANN model taking as input the wind speed of turbine T6. The result is that
the mean absolute error and root mean square error are 20% lower using the model of Section 3. This result is reasonable: due to the severe complexity of the terrain, it is more stable to use the power output of a nearby wind turbine, rather than the nacelle wind speed, as a reference.

In the following, the procedure is reported for estimating the energy improvement. The residuals between the measurement $y$ and the simulation $\hat{y}$, for the data set D1 and D2, are studied. In particular, the interest is in how the residuals vary after the upgrade with respect to before. Therefore, consider Equation 1 with $i = 1, 2$.

$$R(x_i) = y(x_i) - \hat{y}(x_i).$$

For $i = 1, 2$, one computes

$$\Delta_i = 100 \times \frac{\sum_{x \in Data_i} (y(x) - \hat{y}(x))}{\sum_{x \in Data_i} y(x)}$$

Since $\Delta_i$ is constructed with the relative discrepancies of power data each having the same sampling time (10 minutes), the quantity $\Delta = \Delta_2 - \Delta_1$ provides a percentage estimate also of the energy improvement. Further, a Student’s $t$-test can be performed to detect the difference in the residuals $R(x_1)$ and $R(x_2)$. The $t$ statistic is computed as

$$t = \frac{\bar{R}_2 - \bar{R}_1}{\sigma_R \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}.$$  

In Equation 3, $N_1$ and $N_2$ are the number of measurements respectively in D1 and D2, $\bar{R}_2$ and $\bar{R}_1$ are the average residuals between measurement and model respectively in D1 and D2 and $\sigma_R$ is given by

$$\sigma_R = \sqrt{\frac{(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{N_1 + N_2 - 2}},$$

where $S_1$ and $S_2$ are the standard deviations of the residuals in data sets D1 and D2.

The above procedure has been repeated several times with several random choices of D0 and D1 for training and validating the model. Doing this, it is possible to obtain an average estimate of the energy improvement, as well as the standard deviation (and therefore reasonable lower and upper limits for the energy improvement). In Figure 3, for a sample selection of D0 and D1, the sets $R(x_1)$ and $R(x_2)$ are plotted after being averaged within intervals having amplitude of the 10% of the rated power. The order of each bin population is hundreds, for both the data sets $R(x_1)$ and $R(x_2)$. The reported uncertainty bars are the standard deviations of the sub sets inside each interval. Observing Figure 3, it arises that the performance improvement is clearer approaching rated power.
Figure 3. The average difference between measurements and simulation (Equation 1), for data sets D1 and D2 and for a sample run of the model.

The average energy improvement is computed as $\Delta = 3.9\%$. In other words, the estimate is that T7 has produced, during data set D2 and below rated power, the 3.9% more than it would without aerodynamic improvement. The standard deviation is computed as $\sigma_\Delta = 0.4\%$: therefore, reasonable upper and lower limits of the energy improvement are $\Delta_+ = 4.3\%$ and $\Delta_- = 3.5\%$. Further, the probability that there hasn’t been an improvement is computed, using the $t$ statistic, as being of the order of $10^{-14}$.

Moreover, basing on the above result that the average energy improvement below rated power is $\Delta = 3.9\%$, it is possible to estimate how much this amounts with respect to the total energy produced or producible by T7. This has been estimated on the D2 data set and it amounts to $\Delta_E = 1.9\%$. Upper and lower limits are computed to be $\Delta_{E+} = 2.2\%$ and $\Delta_{E-} = 1.8\%$. Further, the estimate has been projected on two 12-months data sets of T7 before upgrade installation and the estimate is that T7 would increase its AEP (Annual Energy Production) of a quantity of the order of $\Delta_{AEP} = 2.0\%$. It is important to notice that this estimate is of the same order of magnitude, but non negligibly lower, than the estimate provided by the wind turbine manufacturer under the hypothesis of ideal operating conditions, i.e. basically when there are no wake-terrain interactions.

5. Conclusions and further directions
This work has been devoted to the objective of quantifying the energy production improvement of a wind turbine whose blades have been retrofitted through the installation of vortex generators and passive flow devices. This wind turbine (T7 in Figure 1) has been the pilot test for the project of retrofitting the blades of all the wind turbines in the wind farm: for this reason, quantifying precisely the energy improvement of turbine T7 in real operation conditions is important for industrial needs. This problem, and in general the issue of quantifying pros and cons of wind turbine technology upgrades, is as well scientifically relevant and therefore in the latest years it has been attracting a certain attention in the literature [3, 6]. Due to the non-stationary conditions to which a wind turbine is subjected, the quantification of an upgrade is a non-trivial problem. In general, two elements are needed: operational data before and after the upgrade and a model for simulating how much the wind turbine would have produced, in the post-upgrade period, if the upgrade had not taken place. The test case of this work was particularly complex as regards the identification of a suitable model: actually, the nacelle wind speed measurements
of T7 could not be used because they are not reliable after the upgrade. Further, the terrain is very complex and the use of a nearby wind turbine as reference of wind speed is controversial. It has been observed that, given the SCADA data sets at disposal, the most precise model for simulating the power production of T7 is constructed with the powers of the nearby wind turbine as inputs. The results for the energy improvement are finally collected in Section 4: the main finding is that the average energy improvement for T7 is estimated as \( \Delta = 3.9\% \) of the energy produced below rated power. This amounts to a \( \Delta_E = 1.9\% \) improvement of the total energy in the first four months (D2 data set) of operation of the retrofitted blades. It is estimated that this amounts to \( \Delta_{AEP} = 2.0\% \) of improvement on the AEP of T7. It is worthy to notice that this estimate has been obtained without filtering out the measurements corresponding to T7 or any nearby wind turbine operating under wake. This is a crucial point for obtaining a realistic computation of the actual improvement because ideal conditions, that commonly are assumed for this kind of estimate, can be very different from real ones. Despite the different methodology and the different hypothesis, the results reported in this work are of the same order of magnitude (albeit not negligibly lower) of the ones provided by the wind turbine manufacturer. The possible further directions of this work are several. The most straightforward future work is repeating the analysis, once more data shall be available for T7 operating with the retrofitted blades: this will surely improve the statical soundness of the results. Further, one of the objectives under development is the numerical and experimental analysis of the mechanical behavior of the upgraded wind turbine, in order to assess how much the loads change in virtue of the retrofitting and if this has some impact on the estimated lifetime of the blades. Another very interesting future direction is extending this kind of analysis to other wind turbines in this wind farm or to other test cases in other terrains. This will allow to study how the energy improvement depends on the terrain and on wake interactions. Finally, the collaboration between the University of Perugia and Renvico srl is devoting to the general issue of wind turbine power curve upgrades. Suitable methods are at study for the assessment of profitability and downsides of different kinds of retrofitting, including control system upgrades (pitch optimization, for example) and extension of the power curve in the very high wind regime.

Appendix A. A crosscheck of the results
As a crosscheck of the results for turbine T7, consider applying the same kind of procedure as above for a wind turbine that hasn’t been retrofitted and observe what changes. T8 has been selected and its power production is the output \( y \) of the model. The inputs \( \mathbf{x} \) of the model are selected to be the powers of T9, T10, T11, T12 and T13 wind turbines.
Figure A1. The average difference between measurements and simulation (Equation 1), for data sets D1 and D2. Wind turbine T8.

In Figure A1, the sets $R(x_1)$ and $R(x_2)$ are plotted after being averaged within intervals having amplitude of the 10% of the rated power. It is the same kind of plot as Figure 3 for wind turbine T7. From Figure A1, it arises that the in the two data sets the discrepancy between measurements and simulations doesn’t change appreciably, differently from what happens for turbine T7. Further, $\Delta$ is computed to be 0.06%.

Bibliography
[1] Wenxian Yang, Peter J Tavner, Christopher J Crabtree, Y Feng, and Y Qiu. Wind turbine condition monitoring: technical and commercial challenges. *Wind Energy*, 17(5):673–693, 2014.
[2] Pierre Tchakoua, René Wamkeue, Mohand Ouhrouche, Fouad Slaoûi-Hasnaoui, Tommy Andy Tameghe, and Gabriel Ekemb. Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges. *Energies*, 7(4):2595–2630, 2014.
[3] Giwhyun Lee, Yu Ding, Le Xie, and Marc G Genton. A kernel plus method for quantifying wind turbine performance upgrades. *Wind Energy*, 18(7):1207–1219, 2015.
[4] László Horváth, Tonći Panza, and Nikola Karadža. The influence of high wind hysteresis effect on wind turbine power production at bura-dominated site. In *Proc. Eur. Wind Energy Conf. Exhibition*, 2007.
[5] Mate Jelavić, Vlaho Petrović, Mario Barišić, and Ivana Ivanović. Wind turbine control beyond the cut-out wind speed. In *Annual Conference and Exhibition of European Wind Energy Association (EWEA2013)*, 2013.
[6] Hoon Hwangbo, Yu Ding, Oliver Eisele, Guido Weinzierl, Ulrich Lang, and Georgios Pechlivanoglou. Quantifying the effect of vortex generator installation on wind power production: An academia-industry case study. *Renewable Energy*, 113:1589–1597, 2017.
[7] Kurt S Hansen, Rebecca J Barthelmie, Leo E Jensen, and Anders Sommer. The impact of turbulence intensity and atmospheric stability on power deficits due to wind turbine wakes at horns rev wind farm. *Wind Energy*, 15(1):183–196, 2012.
[8] Said El-Asha, Lu Zhan, and Giacomo Valerio Lungo. Quantification of power losses due to wind turbine wake interactions through scada, meteorological and wind lidar data. *Wind Energy*, 20(11):1823–1839, 2017.
[9] Po Wen Cheng. Onshore wind energy. *Transition to Renewable Energy Systems*, pages 241–264, 2013.
[10] Francesco Castellani, Davide Astolfi, Massimiliano Burlando, and Ludovico Terzi. Numerical modelling for wind farm operational assessment in complex terrain. *Journal of Wind Engineering and Industrial Aerodynamics*, 147:320–329, 2015.
[11] Francesco Castellani, Davide Astolfi, Matteo Mana, Emanuele Piccioni, Matteo Becchetti, and Ludovico Terzi. Investigation of terrain and wake effects on the performance of wind farms in complex terrain using numerical and experimental data. *Wind Energy*, 20(7):1277–1289, 2017.
[12] Davide Astolfi, Francesco Castellani, and Ludovico Terzi. A study of wind turbine wakes in complex terrain through rans simulation and scada data. *Journal of Solar Energy Engineering*, 140(3):031001, 2018.
[13] IEC. Power performance measurements of electricity producing wind turbines. Technical Report 61400–12, International Electrotechnical Commission, Geneva, Switzerland, 2005.