Applied Energy 185 (2017) 1076–1086

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Analyzing wind turbine directional behavior: SCADA data mining techniques for efficiency and power assessment ☆

Francesco Castellani [a,*], Davide Astolfi [a], Paolo Sdringola [a], Stefania Proietti [b], Ludovico Terzi [c]

[a] Department of Engineering, University of Perugia, Via Duranti, Perugia 06125, Italy
[b] DMII, Guglielmo Marconi University, Via Plinio 44, Roma 00193, Italy
[c] Renvico srl, Via San Gregorio 34, Milano 20124, Italy

HIGHLIGHTS

- The directional behavior of four turbines of an onshore wind farm is investigated.
- The positions of the nacelles are discretized to highlight clusterization effects.
- The recurrent alignment patterns of the cluster are individuated and analyzed.
- The patterns are studied by the point of view of efficiency and power output.
- Significative performance deviations arise among the most frequent configurations.

ARTICLE INFO

Article history:
Received 22 July 2015
Received in revised form 11 December 2015
Accepted 13 December 2015
Available online 28 December 2015

Keywords:
Wind energy
Wind turbines
SCADA control system
Performance evaluation

ABSTRACT

SCADA control systems are the keystone for reliable performance optimization of wind farms. Processing into knowledge the amount of information they spread is a challenging task, involving engineering, physics, statistics and computer science skills. This work deals with SCADA data analysis methods for assessing the importance of how wind turbines align in patterns to the wind direction. In particular it deals with the most common collective phenomenon causing clusters of turbines behaving as a whole, rather than as a collection of individuality: wake effects. The approach is based on the discretization of nacelle position measurements and subsequent post-processing through simple statistical methods. A cluster, severely affected by wakes, from an onshore wind farm, is selected as test case. The dominant alignment patterns of the cluster are identified and analyzed by the point of view of power output and efficiency. It is shown that non-trivial alignments with respect to the wind direction arise and important performance deviations occur among the most frequent configurations.

1. Introduction and motivation

The attention on performance optimization of operating wind farms has recently increased, due to the financial crisis discouraging new investments and lowering energy consumption, and due to the increase of the percentage of energy harvested from stochastic renewable sources. Actually, in order to build smart grids [1], judicious balance of electricity coming from renewable and not renewable sources should be estimated: accurate power forecast and performance assessment are therefore precious. Modern wind turbines are therefore equipped with sophisticated Supervisory Control And Data Acquisition (SCADA) control systems, spreading on 10 min time basis a vast amount of information: details on the wind flow and meteorological conditions, on turbine alignment to the wind, on the conversion of wind kinetic energy into active power, on the vibrational and mechanical status of the machine, on thermal conditions at relevant parts of the turbines, and so on. The scientific and technological challenge is therefore tackling the complexity of SCADA data stream, processing it into novel knowledge and possibly integrating it in the control system itself. Due to the non-trivial physical properties of the source, to the complexity of the machines and of the data flow, the focus of the study spreads the fields of mechanical engineering, physics, statistics and computer science.

☆ This paper was presented at the 7th International Conference on Applied Energy (ICAE2015), March 28–31, 2015, Abu Dhabi, UAE (Original paper title: "How wind turbines alignment to wind direction affects efficiency? A case study through SCADA data mining" and Paper No. 492).

* Corresponding author.
E-mail address: francesco.castellani@unipg.it (F. Castellani).

http://dx.doi.org/10.1016/j.apenergy.2015.12.049
0306-2619/

Machine state dynamics has a typical time scale much more shorter than the SCADA 10-min basis, and its analysis therefore requires a much wider bandwidth. Actually, by the point of view of condition monitoring, abnormal SCADA data measurements can be used only as late-stage indication of a fault. Early fault diagnosis and condition monitoring techniques therefore usually exploit more sophisticated approaches with respect to SCADA data analysis. An exhaustive survey on the state of art on condition monitoring, fault detection and forecast techniques, is provided in [2]. In [3] the most common condition monitoring techniques are listed, moving from their motivations. Gearbox failures are a very common mechanical breakdown in wind energy technology. They impact with a high cost and long downtime, and for this reason the most widespread condition monitoring technique is vibrational analysis for gearbox fault prevention. Actually in [4] it is estimated that a sudden failure of a 1.5 MW wind turbine during winter time leads to around €50,000 of missed production. This amount is up to 5 times greater than the missed production due to a wisely planned maintenance program. This sheds a powerful light on the economic impact of condition monitoring for wind turbine fault prevention. In [3] a comprehensive list is provided of condition monitoring techniques (thermocouple, oil particle counter, ultrasonic testing, vibro-acoustic measurement, torsional vibration, thermography, and so on), their cost, their pro's and con's, the components which are overseen by each of them. Condition monitoring techniques, especially as regards vibration signals, exploit the high capability of Artificial Neural Network (ANN) in pattern recognition from the complexity, which helps in classifying states of the machine. Even though SCADA-based condition monitoring nowadays cannot replace ad-hoc techniques, impressive achievements are being reached. In [5,6] an Adaptive Neuro-Fuzzy Interference Systems (ANFIS) is proposed for monitoring wind turbine SCADA signals.

Numerical models and statistical techniques come at hand not only for SCADA-driven condition monitoring, but also for optimization of operating wind farms. In [7] a review is provided on performance optimization, summarizing objectives, targets and formulations. Machine learning from SCADA data is a rapidly growing subject: in [8] the most fundamental wind turbine performance characteristic curve, i.e. power curve, is reinterpreted through data-driven models, which can be used as a reference profile for on-line monitoring of the power generation process. In [9], data-mining algorithms are used to construct prediction models for wind turbine faults. Random forest algorithm models are shown to provide the best accuracy among all algorithms tested. The robustness of the predictive model is validated against experimental data. In [10] four data-mining approaches for wind turbine power curve monitoring are compared and sensible impact of ambient temperature and wind direction is highlighted. In [11] pitch faults are analyzed. Via an inductive rule learner, it has been possible to separate turbine states in three classes with 85% accuracy: no pitch fault, potential pitch fault, pitch fault established. From these references and many more, it arises that classifying data in categories is fundamental in order to prevent faults or evaluate performances through machine-learning models. This is commonly possible through a blind approach, within a certain extent, considering one turbine at a time, as in the examples above.

Yet, there is a typical phenomenon which involves more wind turbines at the same time: wakes. They are the most common cause of severe power losses during the operative phase and have therefore stimulated a vast amount of scientific literature. Horns Rev wind farm in Denmark has been the object of several studies about wakes [12,13]. In [14] for the Horns Rev test case, a systematic analysis is carried in order to highlight dependency of power deficit on wind rose, wind speed, turbulence intensity and stability of the atmosphere. Peculiar attention has been devoted to wakes, as interpreted by the point of view of wind direction [15] and yaw alignment to it: in [16] turbine nacelle misalignment under a range of downstream wake angles is highlighted, indicating a characteristic of wind turbine behavior not generally considered in wake studies. It is shown that the turbines yaw independently in order to capture the increased wind speeds present due to the lateral influx of turbulent wind, contrary to many experimental and simulation methods found in the literature. On the grounds above, suggestive hints arise that, for interpreting wakes, it would be valuable to upgrade from the analysis of wind direction to the analysis of nacelle configurations as a response to wind direction. Identifying and classifying configurations of turbines under wake therefore should take into account wind direction, yaw response to it, and possibly clusterization effect. Actually first glimpses of wake characterization through clusterization are collected in [17,18]. The test case is the same as in the present work: a cluster of turbines of an onshore wind farm, affected by multiple wakes when the wind blows from the 270° sector. It is shown that, when severe wakes manifest, the cluster tends to behave as a whole rather than as a collection of individuality and considerable power losses occur. The lesson, which is collected and developed in the present work, is that an underworld of nacelle and cluster configurations exists, associated to each wind condition.

The motivation of the present work therefore is the difficulty of discriminating the quality of turbine states, by on operational point of view, through blind data-driven models when a collective phenomenon as wakes dominates the dynamics. Yet, patterns do exist and must be recognized. If blind models cannot be effective, a first input to the subject might be provided by human boost, through the knowledge of wake phenomenon. This is exactly where the present work lies its foundation. The wind farm is subject to a global phenomenon, the wind flow, which might suffer very local effects. Further, each turbine has its own independent control system, which could have different responses resulting in slight different and unique yawing position distributions. On these grounds, the philosophy of this study is that the patterns to identify and classify should, because of the complexity and the global nature of the phenomenon, be physical: the patterns we propose are indeed the nacelle alignment configurations of the cluster to the wind. This is supported, apart from intuitiveness, by the fact that wind direction and response of the yawing system to it are highlighted as crucial agents of wind turbines when there are considerable wakes [15,16]. A cluster of turbines, lying very close to each other and affected by severe wakes along the most common wind direction, is employed as testing ground. A SCADA post-processing method is proposed for identifying and separating alignment patterns of the cluster: nacelle positions are discretized with a judicious grain and it shown that this automatically leads to the identification of patterns. Subsequently, tools are to be developed in order to discriminate patterns: some are proposed in Section 3. For example, the number of times the pattern occur and the characteristic time distance between one occurrence and the subsequent, the average power output of the cluster, the efficiency of the cluster. Efficiency deserves a little discussion: in [19] it is shown that efficiency cannot be defined as usual in the literature, even for a test case sited on a very gentle terrain, which is exactly the same of this work. Therefore, in the present work, the alignment configurations for the selected cluster are also classified according to the efficiency, as redefined in [19], in order to make it consistent for the onshore case. The main conclusion of this study is that the discretization approach indeed provides a consistent pattern separation and identification, and the patterns might be classified and evaluated intuitively through the formulated indicators. This constitutes somehow a first analysis of alignment patterns of clusters of turbines through SCADA data-mining, which would be valuable to encode in more sophisticated models or to cross with more complex sources of information. The future perspective is that, learning from the knowledge that each farm has of itself through historical SCADA data sets, custom alignment rules to the wind direction could be favorable and might become realizable: this is the philosophy of yaw active control systems, which are recently attracting a considerable attention in the scientific literature [20,21].

The structure of the paper is as follows: in Section 2 the test case wind farm and the testing ground (selected cluster, meteorological regime) are described. In Section 3, the SCADA data mining method and the analysis are introduced. In Section 4, the nacelle configurations are analyzed by the point of view of their recurrence (frequency of recurrence, time distance between one recurrence and the subsequent), power, efficiency. Further, some statistical Indexes, introduced in Section 3, are discussed. It is shown that sensible performance deviations arise between the most frequent patterns of the cluster and that a consistent approach needs to take into account both efficiency and average power output associated to a given configuration. Conclusions and further directions of this work are drawn in Section 5.

2. The wind farm

The analysis is performed on the test case of a wind farm sited in southern Italy. The machines have 2 MW of rated power, 82 meters of rotor diameter and 80 meters of hub height. The main features are summarized in Table 1.

The distance between the closest turbines is of the order of two rotor diameters. From the layout of Fig. 1 it therefore arises that there are considerable wake interactions when wind, having moderate intensity, blows from West.

The wind rose of Fig. 2 reveals that this indeed happens very frequently, and this allows to have at hand a vast data set for the analysis of the present work.

Further, the farm lies on a very gentle terrain and it therefore is a hyphen between the features of offshore wind farms and onshore wind farms, where commonly the effects of orography are non-negligible and difficult to disentangle from those of wakes.

The inter-turbine distance and the population of the wind rose (Fig. 2) allow several possible choices of subclusters to analyze, as for example T40-T56-T57. As shown in [17], when the wind blows from 270°, the most under-performing turbine of the wind farm is T58. For this reason, a cluster including T58 is selected: the one from T57 to T60. Further, in order to apply the method discussed in Section 3, it is precious to have a reliable milestone for wind flow conditions. The turbines of the cluster lie symmetrically to the met-mast and close to it. This is another key point in supporting the choice of T57–T60 cluster for the analysis. When the wind blows from West, the met-mast is slightly disturbed by turbine T55. Yet, the distance between T55 and met-mast is 12 rotor diameters and this guarantees that wakes from T55 at met-mast are strongly mitigated [22]. In any case, in Section 3, a simple method to remove possible bias of wind direction measurements at met-mast (due to wakes or incorrect calibration) is proposed. The same consideration about wake recovery applies to turbine T57, which is influenced by turbines T53 and T55 when the wind blows from West: the distance between T57 and the T53–T55 subcluster is 10 rotor diameters and so the selected cluster T57–T60 can be considered to also have at its upstream turbine a reliable approximation of free flow when the wind blows from 270°.

Table 1
Main features of the test case wind farm.

Feature	Value
Number of turbines	9
Rotor diameter (m)	82
Hub height (m)	80
Rated power (MW)	2
Terrain type	Flat

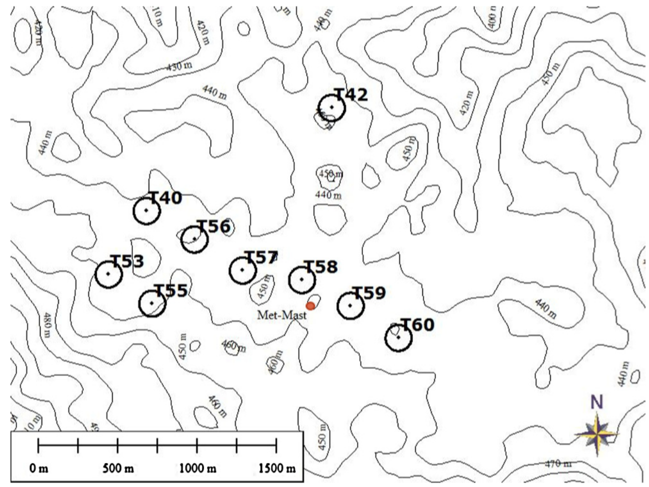


Fig. 1. Wind farm 1 layout with met-mast position.

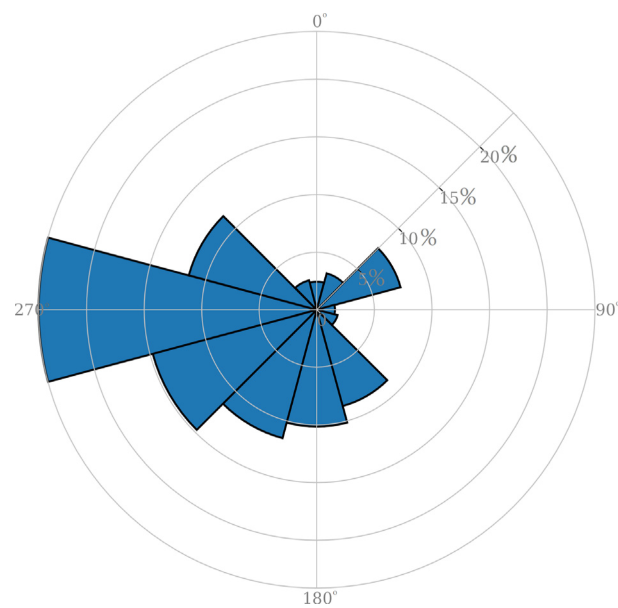


Fig. 2. Wind rose. Wind direction (as measured by met-mast) frequency percentages, on 12 sectors basis.

3. The method

A data set, built of two years of meteorological data (wind direction and intensity at met-mast at hub height), turbine power output and nacelle positions is collected. The met-mast wind vane is calibrated by computing the relative speed-up against the nearest T58 and T59 turbines as a function of wind direction, and averaging on narrow wind direction bins. On such a flat terrain, the extremal speed-up is expected to occur along the direction of the geometric line connecting the met-mast to the nearby turbines. With this procedure, a 4° offset is highlighted. Data are subsequently filtered as follows:

Table 2
Example of nacelle position codifying: most frequent cluster configuration.

Turbine	Nacelle position (°)	Binary coding
T57	272	000000010000000
T58	260	000010000000000
T59	264	000001000000000
T60	256	000100000000000

- They are synchronized and filtered on the requirement of simultaneous power output production from turbines T57 to T60. Applying this filter, the number of 10-min measurements reduces to 58,886.
- They are filtered on the requirement that wind speed at met-mast is between 4 and 8 m/s. This range is selected for several reasons: first, because the wind rose is populated mainly by moderately low wind speeds in the 270° sector. This matches with the intent of analyzing efficiency and power response of aligning configurations in the most challenging regime: when the thrust is high, wake effects are relevant and the meandering of the wind is typically more severe and induces the difficulty of downstream turbines in following the wind direction. The amplitude is selected also because it is of the order of magnitude of two typical standard deviations on a wind speed measurement.
- They are filtered according to wind direction as measured by the met-mast at hub height: the selected interval is $275^\circ \pm 5^\circ$. This choice takes into account the systematic bias on wind direction measurements discussed above, and should correspond to an actual wind flow from 270°. The amplitude of the interval corresponds to two typical standard deviations on wind direction measurements.
- Outliers due to manifestly incorrect alignments of the turbines, i.e. outside a 60° range around 270°, are filtered away.

The following objective is discretizing nacelle positions on a grain fine enough to appreciate the differences in alignment patterns, coarse enough to provide a certain clustering of the measured records. At this aim, an estimate is provided on the movement of the nacelles from a given 10 min time step to the one immediately following. The distribution of such movements is analyzed and it is noticed that, in more than the 30% of the cases, a movement greater or equal than 5° is observed. Therefore, a 4° grain is selected. Some interesting justification on this choice is also provided in Section 4. A bit is associated to each of the 15 elementary 4° intervals of wind direction amplitude ranging in the $[240^\circ, 300^\circ]$ interval. Each bit turns from 0 to 1 if the nacelle position measurement falls inside it. For each turbine of the cluster, for each 10 min time step, a binary number, made of 15 digits only of which can be 1 (a measurement cannot fall in two interval simultaneously), is therefore associated. The numbers associated to each turbine are glued together with the correct ordering (from T57 to T60) and this gives a unique 60-bit binary number for the cluster for each time step. An example is provided in Table 2, where the most frequent configuration of the cluster is represented in the binary coding. In the second column of Table 2, the nacelle position rounded to 4° precision (as the grain) is shown, in the third column the corresponding coding is shown: for example, 272° corresponds to the eighth out of fifteen 4° interval in the $[240^\circ, 300^\circ]$ window and therefore the corresponding binary number is a sequence of seven 0's, a 1, seven 0's. Similarly for the other single turbine entries. The unique number for the cluster is obtained just gluing together the entries of Table 2: 00000001000000000001000 0000000000001000000000000100000000000.

The total number of possible configurations for describing the cluster is therefore $15^4 = 50{,}625$. This number provides a reasonably vast underworld of patterns for describing the alignment behavior of the cluster, also because each of the single turbine data set of nacelle measurements actually spans all the $[240^\circ, 300^\circ]$, except for T60 which never records occupation of the first bin out of fifteen. The next step is classifying the configurations according to the frequency of their occurrence and identifying the dominant patterns of the cluster: regrouping measurements according to the configuration to which they are associated, one starts from a data set and obtains N data sets, where N is the total number of different configurations. The data flow and the operations performed on it are outlined in Fig. 3.

These N data sets can be studied by several points of view:

- *Size*. How many times does a given configuration occur? This is the fundamental information because the aim is investigating if the most frequent alignment patterns are favorable or not and how much. Therefore, configurations are ranked according to their weight in the data set, i.e. frequency of occurrence.
- *Compactness*. It is important to understand if a configuration is episodic and occurs once for a given period of time (possibly for many time steps) and never more, or rather if it is recurring several times in the time history of the cluster. In the latter case, it is more valuable to understand its features. In order to quantify this aspect, an Index is proposed measuring the compactness of the configuration. It is defined as follows, Eq. (1).

$$I_{\mathrm{Com}} = 100\frac{N_{\mathrm{adj.rec.}}}{N_{\mathrm{rec.}}} \quad (1)$$

To quantify compactness, we examine the number $N_{\mathrm{rec.}}$ of records of which the i-th configuration data set is built. How many of these are consecutive to the previous? $N_{\mathrm{adj.rec.}}$ is therefore the number of consecutive records of the i-th configuration data set. It is reasonable to let a certain gauge, which means considering as belonging to the same time block two records not strictly consecutive, but separated by a reasonably low time interval: that is what we call almost consecutive records. The selected time threshold for considering two time steps

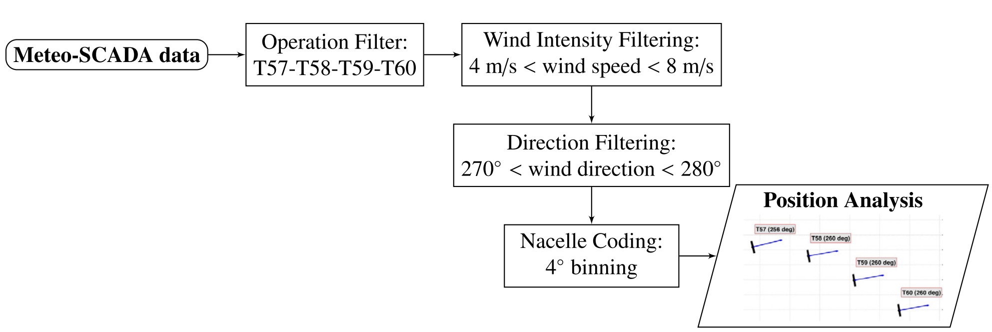
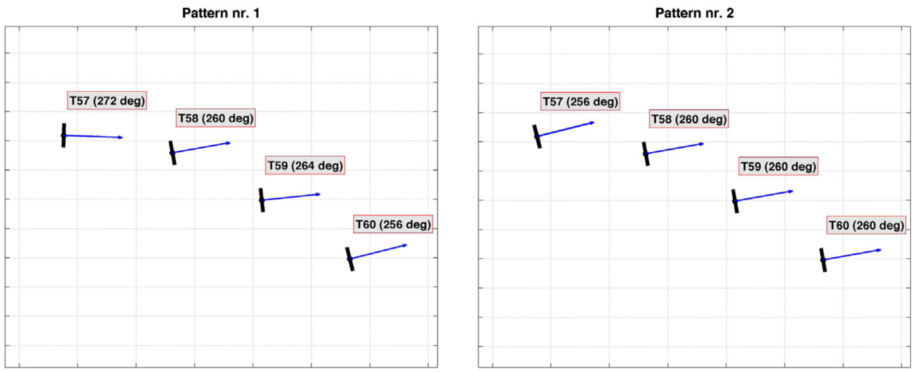
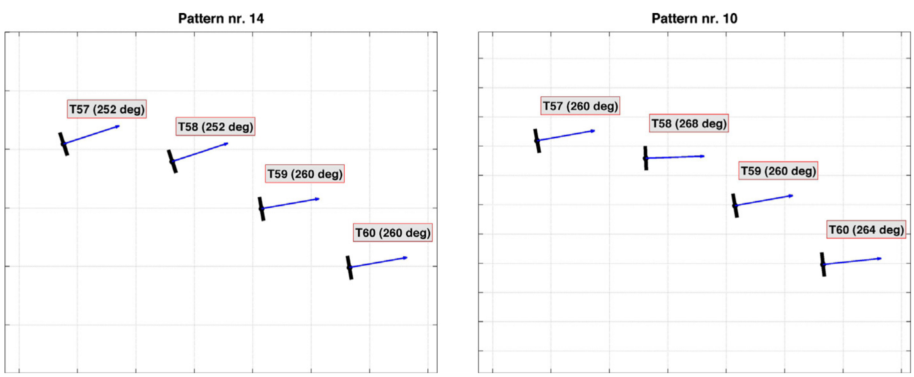


Fig. 3. Data flow for the postprocessing.



(a) First most frequent configuration (b) Second most frequent configuration

Fig. 4. The two most frequent configurations of the cluster.



(a) First most productive configuration (b) Second most productive configuration

Fig. 5. The two most productive configurations of the cluster.

almost consecutive is one hour. Summarizing, I_{Com} tending to 100 means that a configuration is observed for a very compact period and it does not occur anymore, while instead low values of I_{Com} indicate that a configuration does not occur compactly in time. Another useful metric for quantifying compactness of a given configuration is the maximum time interval between two subsequent records associated to the i-th configuration.

- *Efficiency*. As discussed in Section 1, efficiency is one of the main metrics for assessing operational behavior of a farm or a cluster, mainly because it is intuitive and easy to compute and simulate numerically. In [23,19] it is shown that, onshore, efficiency cannot be defined as usual in the literature because the effects of the terrain make the approach not consistent. A novel definition is proposed and discussed and one of the test cases is exactly the wind farm analyzed in the present study. The same definition of efficiency is employed here on. The efficiency can be computed on 10 min time basis for a group of N turbines ($N = 4$ for the case of the T57–T60 subcluster). It reads as in Eq. (2), where P_i is the 10-min average of power output of the i-th turbine of the cluster. Basically, Eq. (2) defines efficiency on 10-min time basis as the average cluster power in units of the power of the best performing turbine of the cluster.

$$\eta = \frac{\sum_{i=1}^{N} P_i}{N \cdot \max(P_i)} \tag{2}$$

For each of the records of the data set associated to the i-th configuration, the efficiency of the cluster can be computed. These values can be averaged on the data set, obtaining one measure of mean efficiency $\hat{\eta}$ for each cluster configuration. Further, an Efficiency Index can be defined for each configuration as follows, Eq. (3).

$$I_\eta = 100\left(\frac{\hat{\eta}}{\bar{\eta}} - 1\right) \tag{3}$$

In Eq. (3), $\bar{\eta}$ is the average efficiency of the cluster on the whole 270° data set and $\hat{\eta}$, as above, is the mean efficiency of a configuration. I_η therefore quantifies how much the mean cluster efficiency deviates (in good or bad) from the average efficiency under the given meteorological conditions.

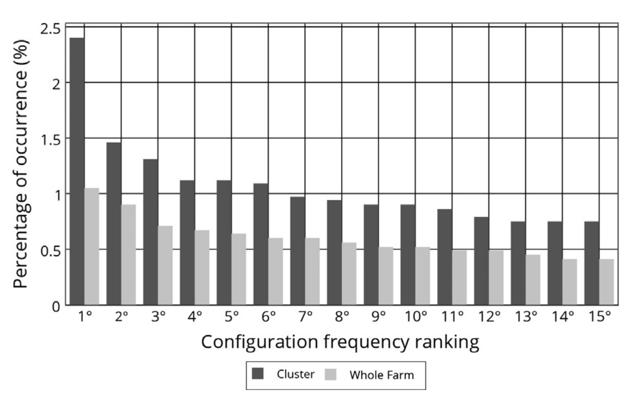


Fig. 6. The weight of the fifteen most frequent cluster and wind farm configurations.

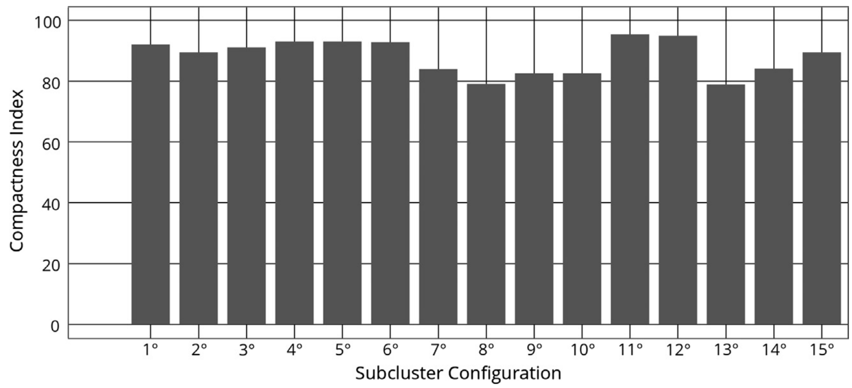


Fig. 7. Compactness index for the fifteen most frequent configurations of the T57-T60 subcluster.

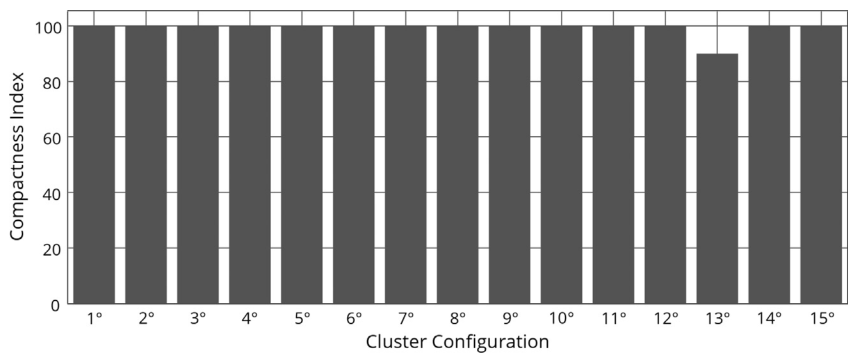


Fig. 8. Compactness index for the fifteen most frequent configurations of the wind farm.

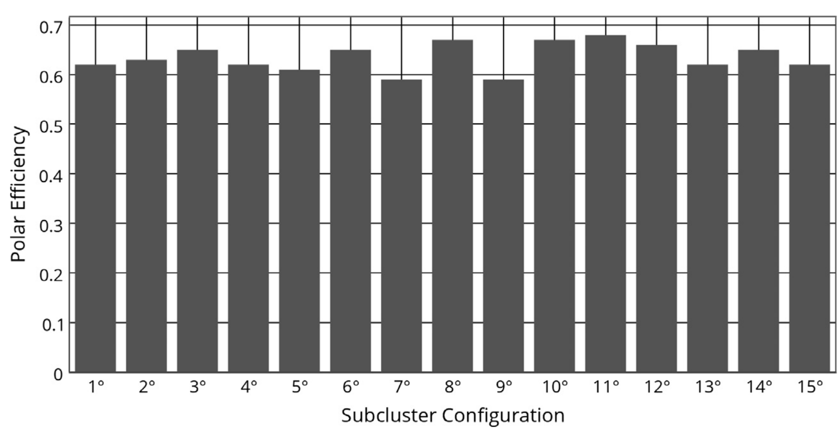


Fig. 9. Average efficiency for the fifteen most frequent configurations of subcluster.

- *Power output*. For each of the records associated to each configuration, the mean power of the cluster can be computed. These mean powers can subsequently be averaged, obtaining a single power estimate per configuration. Further, a Power Index is defined in Eq. (4).

$$\beta = 100\frac{N_{\mathrm{ab.mean}}}{N_{\mathrm{tot}}} \tag{4}$$

Given the data set, whose size is N_{tot}, consisting of all the records passing the filters described above, $N_{\mathrm{ab.mean}}$ is the number of times that the single power estimate (defined here above) of the i-th configuration is greater than the average cluster power. Therefore, for each configuration, two estimates are provided on their quality by the point of view of power output: the average configuration power and the Power Index of Eq. (4).

The same techniques are applied also to the whole wind farm producing output in unison, because this provides further insight on the reliability and the limits of the method. It actually rather is an absurdum case study, which is advocated expressly because it demonstrates that the methods collapse and are not able to codify meaningful information for such a vast number of machines on such vast terrain. In other words, when wake effects are relevant, a compact cluster as T57–T60, at least in the sense codified by our approach, can be said to behave as a whole rather than as a collection of individuality. The same conclusion instead cannot be drawn

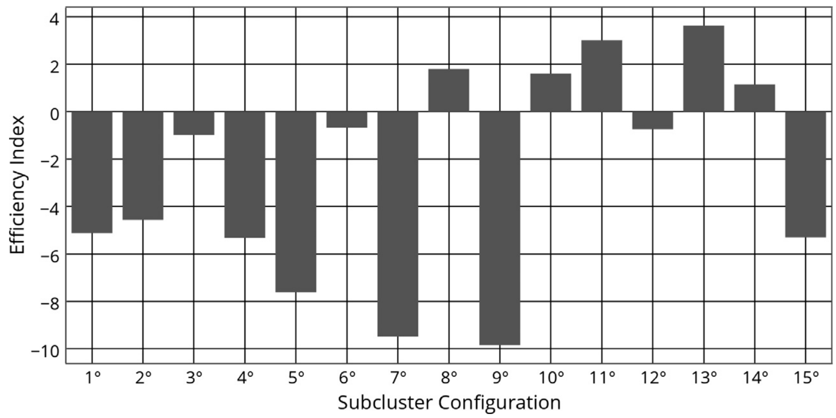


Fig. 10. Efficiency index for the fifteen most frequent configurations of subcluster.

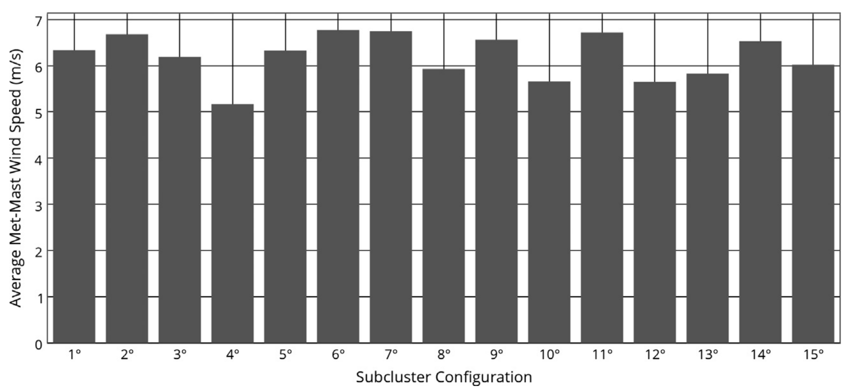


Fig. 11. Average met-mast wind speed for the fifteen most frequent configurations of subcluster.

for the whole farm. Results about the whole wind farm are therefore limited and reported as useful for this kind of discussion.

4. Results

In this section, results are displayed for the fifteen most frequent configurations of the subcluster. They are collected according to the methods proposed in Section 3. First, some general comments arise from analyzing the relative positioning of the turbines. For brevity, in Fig. 4 the two most frequent configurations are displayed, and in Fig. 5 the two most powerful configurations are displayed. They share a common ground, which is a northward wake distortion, with respect to the undisturbed inflow, of the positioning of the T58–T60 turbines. This is reasonable, because a parallel, northward wake distorting alignment is expected to reduce the effect of wakes. Yet, as discussed later on, non-trivial performance deviations arise among the configurations.

Here on, some meaningful results are shown, parallel to the methods proposed in Section 3. For all the figures from 6–13, the configurations are labeled in the x-axis with a progressive number indicating their order as concerns frequency of occurrence: 1° is the most frequent and so on.

- *Size*. In Fig. 6 the frequencies of occurrence of the fifteen dominant subcluster configurations are reported. Despite a vast amount of total configurations actually and possibly occupied, more than the 15% of the total data set falls in one of the 15 most frequent configurations. As an absurdum test case, the frequencies of the configurations are analyzed also for the whole wind farm. The absolute values of the first frequencies (see Fig. 6), and their much more flat distribution with respect to the case of the cluster, are first hints of the fact that it is not much consistent to talk at all about patterns for a whole wind farm.
- *Compactness*. Fig. 7 displays the Compactness Index, defined in Eq. (1), for the fifteen most frequent configurations. The compactness of the whole data set passing the meteorological filter is 83.40%. Comparing against Fig. 7, it therefore arises that the configurations display a compactness of the same order of magnitude of the parent data set. This means that a not negligible fraction of the records, during which a given configuration manifests, does not occur consecutively (or almost) to the previous one. Actually it is observed that all the most frequent subcluster configurations recur even with a distance of more than 200 days from one appearance to the other. This means that the configurations of the subclusters are not time packets manifesting once, for a certain consecutive time, and never more: they are instead characteristic alignment patterns which recur. This strongly supports the motivation of this study and the reliability of the approach.

Fig. 8 shows the Compactness Index of Eq. (1) for the fifteen most frequent wind farms configurations. It arises that basically all the configurations manifest in almost consecutive time packets. Therefore the configurations cannot indeed be said to be recurrent in time. This proves that it makes sense to treat together a limited number of turbines, as in the case of the T57–T60, especially because their behavior under the 270° regime has a common ground: wake effects. Instead, when a cluster is not a cluster, but a whole and vast wind farm, the proposed indicators, especially I_{Com}, lose their meaning. This is the lesson from Fig. 8, and from the absurdum test case of treating also the whole wind farm as a cluster. It is a useful clue of the consistency of the methods: they describe patterns of compact clusters, and some of their features. The limit $I_{\mathrm{Com}} \to 100$ deserves another observation: if in this limit it does not make sense to speak of patterns, this can be a posteriori used for discriminating the binning for the nacelle discretization procedure. This sensitivity analysis has been performed and it has been observed that using a 3° binning interval, I_{Com} for the most frequent subcluster configurations tends to 100. Therefore the proposed 4° discretization is selected because it is the finer grain providing consistency of the approach.

- *Efficiency*. Fig. 9 displays the mean configuration efficiency $\hat{\eta}$ for the fifteen most frequent patterns of the subcluster and Fig. 10 displays the Efficiency Index of Eq. (3).
  The average efficiency of the subcluster $\bar{\eta}$ is equal to 0.66, with a 0.09 standard deviation. This means that many of the most frequent configurations display an average efficiency lower than the subcluster does in general when the wind blows from the 270° sector. Actually Figure (3) displays that I_η varies from -10% to +4%, i.e. one observes relatively modest variations (less than one measured σ) against the mean 270° efficiency, and most frequently negative.
- *Power*. Fig. 12 shows the average configuration power, normalized to one turbine. The average normalized power, measured on the whole 270° data set, amounts to 270.32 kW, with a 149.21 kW standard deviation. This means at first that, in proportion, power varies much more than efficiency does from a pattern to an other. Further, from Fig. 12 it arises that the most frequent configurations display an average normalized power ranging around the measured cluster mean, with an oscillation of the order of magnitude of the measured σ. The oscillations around the mean are both positive and negative, while in the case of efficiency they are most likely negative. A useful inquiry on the causes of these power deviations is obtained by considering another quantity: the average wind speed, as measured by the met-mast, on each recurrent configuration. This quantity, obtained by simply averaging the wind speed measurements from the met-mast on each subset constituted by the records associated to each configuration, is plotted in Fig. 11. From the Figure it arises that, even though the interval of wind speeds allowed by the data filtering is relatively sloppy window $[4, 8]$ m/s, actually the averages on the recurrent configurations are far more compact and they spread the $[5.2, 6.8]$ m/s wind speed interval. Further, the wind speed averages of Fig. 11 are negligibly correlated to the average powers of Fig. 12. For example, the

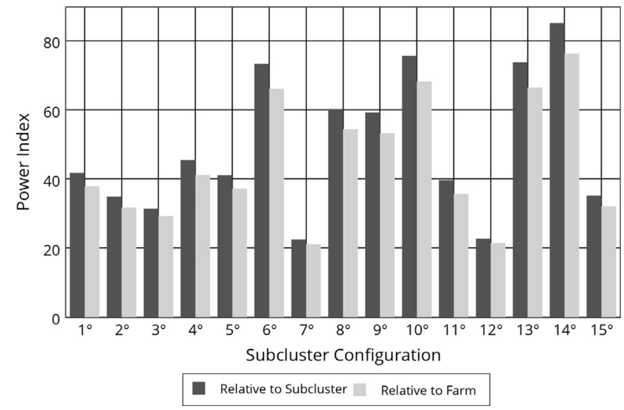


Fig. 13. Power Index for the fifteen most frequent configurations of subcluster: relative to subcluster and relative to wind farm.

eleventh configuration displays the second highest average wind speed, but the average power associated to it is of the order of one half of the best powers recorded in Fig. 12. These facts provide reliable consistency check of the approach and of the selection of the filtering amplitudes. Further, they are a strong clue of the fact that performance variations are indeed not due to external conditions, but rather to response of the cluster to them.

As concerns the relationship with efficiency, Fig. 12 against Fig. 10, a subtlety arises: efficiency is in general a good metric for evaluating the quality of performances, but it can be biased and is not self-sufficient for providing a reliable estimate. Actually, an homogeneous bad quality behavior is not encoded by efficiency as bad (see Eq. (2)). For example the 11° most frequent configuration displays an efficiency higher than the mean, while the power is instead lower. The other cases of high efficiency configurations are instead associated also to higher performances by the point of view of power output. This means that efficiency is a good metric, but a twofold approach (efficiency – power) is needed to assess performances more safely.

Fig. 13 displays the Power Index of Eq. (4). It is computed in two ways: comparing against the normalized cluster power, and against the normalized farm average power. It is expected that an alignment pattern of the subcluster, suffering multiple wakes in the 270° sector, is optimized against the pattern itself, rather than against the whole wind farm, which has also several upstream turbines when the wind blows from West. Therefore,

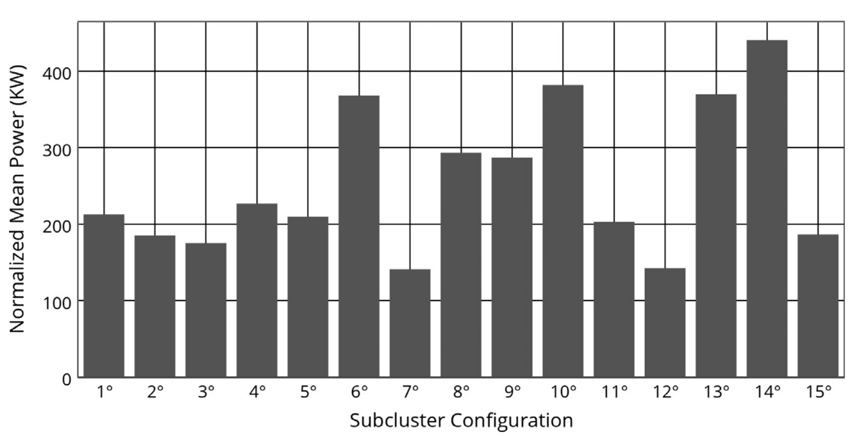


Fig. 12. Normalized average power output for the fifteen most frequent configurations of subcluster.

it is expected that β is higher when computed relative to the subcluster rather than to the farm. Fig. 13 shows that this is indeed the case. In this case, this simply provides a consistency check, but in general the approach allows to highlight manifest and severe misalignment resulting in degraded performances.

5. Conclusions and further directions

This study deals with performance assessment for onshore wind farms. In particular, it deals with the a common phenomenon, involving multiple turbines and characterized by several concurring factors [14]: wakes. Wakes are indeed a "glocal" phenomenon: the wind farm is subject to a global agent, the wind flow, but at each turbine site severe local affects might arise, especially in complex terrains. Further, the control system of each turbine might react differently, resulting in an underworld of yawing positions, each of them characterized by peculiar performances and mechanical stress. The philosophy of this study is attempting a description of a cluster of turbines, severely affected by wakes, as a whole, rather than as a collection of individuality. First developments on the subject have been achieved in [17,18]. The objective is employing SCADA data-mining techniques for automatically identifying patterns of the cluster. Subsequently, some methods for evaluating and discriminating the patterns are proposed and discussed. Actually distinguishing goodness of a state or of a measurement is crucial in data-driven learning techniques: data must be divided, roughly speaking, in the minimum possible number of hyperplanes. Reasonable classifications are good and bad, or good, bad, possibly bad. In particular, the underlying idea of this study is that the patterns to identify and classify should, because of the complexity of the phenomenon, be physical and intuitive as much as possible: and actually the patterns we propose are indeed the nacelle alignment patterns of the cluster to the wind. This looks particularly fit and consistent because wind direction and individual response of the yawing system to it are highlighted as crucial agents of wind turbines performances when there are considerable wakes [15,16].

On these grounds, two operations are performed in this study:

- A SCADA post-processing method is built, which allows to automatically identify the patterns of a subcluster.
- Methods are formulated and discussed for building indicators of the quality of these patterns.

This approach is tested on two years of data of a subcluster of four turbines, from an onshore wind farm sited in southern Italy. The inter-turbine distance and the layout are such that considerable wake effects arise when the wind blows from the 270° sector (very populated in the wind rose, Fig. 2). Further, a met-mast lies quite symmetrically with respect to the cluster and, when the wind blows from West, captures (with fairly good approximation) free wind. The post-processing method is based on the discretization of nacelle position on a judicious fine grain, as discussed in detail in Section 3. A 4° grain is selected: the reasons of this threshold and ex-post considerations about its goodness are addressed respectively in Sections 3 and 4. A bit to each elementary 4° interval of each turbine is associated, turning from 0 to 1 if the corresponding nacelle measurement falls inside the interval itself. Applying other post-processing methods for outliers removal (manifest incorrect nacelle alignment) discussed in Section 3, in the present case one ends up with a 15-bit number to each turbine for each 10 min time step. Gluing the numbers for each turbine of the cluster, a unique 60-bit number is obtained for each 10 min time interval of the cluster selected in this study. Discretization inevitably leads to a finite number of patterns, which are to be somehow classified and evaluated. This is at least partially achieved through the indicators formulated in Section 3. Further, in some cases, an absurdum testing ground is included: the limit in which the cluster becomes the whole wind farm. In Section 4 it is shown that the proposed indicators lose their meaning in this limit. This supports that the proposed methods highlight some features typical of clusterization phenomenon, which are not present when the cluster indeed is not a cluster. The following are some meaningful observations collected from the proposed methods:

- The proposed formulation and identification of patterns indeed is consistent because, under a vast totality of possibilities and configurations actually occupied, some patterns are far more frequent than others. In particular, more than the 15% of the measurements falls in one of the 15 most frequent patterns of the cluster.
- Patterns are indeed recurrent, and not manifesting once for a compact period of time and never more. This is quantified through a Compactness Index formulated in Eq. (1). In particular, this indicator is good for consistency check: when computed for the whole wind farm, it definitely loses sense and it demonstrates that patterns simply do not exist for a whole vast farm. The lesson is that only compact clusters, as the one selected, might be described in the language of recurring patterns.
- To each cluster, two main metrics are associated: normalized mean power and efficiency (as defined in [19]). The lesson is that mean power varies much more than efficiency from one configuration to the other: cluster power is, as intuitively expected, a very responsive parameter for distinguishing patterns and evaluating them. This seems not to be the case for efficiency, or at least not as much. In [19] it is shown that efficiency, even though redefined through Eq. (2) employed also in the present paper, is able to macroscopically capture regimes of underperformances, at least for the gentle terrain case, which is exactly the same of this work. Actually also macroscopically the situation is in general more complex: efficiency does not quantify goodness of performances in an absolute sense, but rather captures the degree of homogeneity in the behavior of a wind farm or cluster of turbines. For offshore wind farms, where power losses and in-homogeneity are mainly due to wakes, efficiency becomes more and more synonymous of good quality. Onshore, this is not expected with such a regularity and in [19], it is shown that in very challenging sites, in-homogeneity due to wind flow acceleration induced by the terrain might improve performances of clusters of wind turbines. The analysis of this work show that, even though the selected test case represents somehow a halfway between the features of an onshore and offshore wind farm, when one ventures in the underworld of nacelle configurations, efficiency is not a totally responsive parameter: microscopically, efficiency must be handled with care. It is therefore reasonable to find out that most of the energetically favorable configurations also display good average efficiency, but this should not be expected as a strict rule. For example the 11° most frequent configuration displays an efficiency higher than the mean, while the power is instead considerably lower.
- On the recurrent configurations, negligible correlation arises between normalized mean power and average met-mast wind speed. In particular, examples of configurations associated to high wind speeds (relative to the filtered data set) and low average performances often occur. This is a strong clue of the fact that the performance deviations observed in this work are not due to changing external conditions, but rather to changing of the response to external conditions. In other words, there is an underworld (nacelles and clusters as codified in this work) to explore further.

- A Power Index is formulated in Eq. (4). It is shown to be a very responsive Index, which might be useful for discriminating quality of the patterns.

In general, the outcomes of this work are interesting for the perspective of improving wind farm operation. Wakes are a very common, yet very stressing phenomenon for a wind turbine: the management is usually set according to the criterion of preserving the turbines from fatigue loads. For example, in [24] the aim of a compromise between power output and fatigue load consumption is clearly highlighted, and a typical too conservative management is cited: the so called Pattern Derating, employed in California, which is basically shutting down very second turbine of a row when the wind blows along its inclination. A stimulating question is nowadays becoming feasible: within what extent can one deviate from this kind of policies? How much is it reasonable to force custom alignment patterns (yaw active control) [20,21]? What are the pro's and what are the con's of a custom policy? This work focuses on the pro's side: it shows that the underworld of nacelle configurations can be tackled and classified in a consistent way and that there are sensible performance deviations among the patterns. This work strongly supports, therefore, that consequently also the second question is worth addressing: what are the con's of a custom alignment policy when there are wakes? This is the main further direction the authors plan to investigate in the future: pairing the experimental SCADA analysis approach of this work with a mechanical analysis. Usefulness of this issue for improving wind farm operation also arises from the discussion in [16], where it is argued that active wind farm management is based on the idea of reducing axial induction by active control of yaw, tip speed ratio, pitch, but mere numerical simulations cannot take into account that a reduction in axial induction can alter lateral and vertical mixing of free-stream wind and therefore the yaw alignment of the downstream turbines. The wind farm selected for this study does not have a condition monitoring system, but it has been chosen because its features (gentle terrain, considerable wake effects) make it very fit for testing the reliability of the approach and the pro's side of the problem. On these grounds, a very stimulating test case is planned to be investigated: a wind farm in complex terrains, as for example the one addressed in [25], which has also been studied as a test case of IEA-Task 31 Wakebench project [26] aiming at developing frameworks for the evaluation of wind farm flow models operating at microscale level. This shall be done for several reasons: because this wind farm has also a condition monitoring system recording vibration signals on the kHz scale, because in this wind farm there are clusters of turbines similar to the one addressed in this work, but characterized by a mixture of wakes and complexity effects, because this wind farm features different turbine models with respect to the test case of the present study (therefore it would be a valuable testing ground to assess the influence of the wind turbine technology on the proposed analysis). Another interesting further direction is linking the approach of this study to numerical models. First developments have been reached in [25] for the complex terrain test case of [26], but in that study basically the operational data are averaged, in order to compare against the steady flow predictions of the solution to Reynolds Averaged Navier Stokes equations, obtained with the WindSim numerical tool [27]. In order to get insight into the dynamic underworld of nacelle alignment of patterns of turbines, more complex numerical approaches are needed, as Large-Eddy Simulations [28] providing a time series of the evolution of turbulence and free wind flow, possibly combined with simulation techniques describing the power extraction of the wind turbine [29,30].

References

[1] Colak I, Fulli G, Sagiroglu S, Yesilbudak M, Covrig C-F. Smart grid projects in Europe: current status, maturity and future scenarios. Appl Energy 2015;152:58–70.

[2] Kusiak A, Zhang Z, Verma A. Prediction, operations, and condition monitoring in wind energy. Energy 2013;60(0):1–12. http://dx.doi.org/10.1016/j.energy.2013.07.051. <http://www.sciencedirect.com/science/article/pii/S0360544213006579>.

[3] Yang W, Tavner PJ, Crabtree CJ, Feng Y, Qiu Y. Wind turbine condition monitoring: technical and commercial challenges. Wind Energy 2014;17(5):673–93.

[4] Wind Power Monthly. Condition monitoring, expert report edition; 2013.

[5] Schlechtingen M, Santos IF, Achiche S. Wind turbine condition monitoring based on SCADA data using normal behavior models. Part 1: System description. Appl Soft Comput 2013;13(1):259–70. http://dx.doi.org/10.1016/j.asoc.2012.08.033. <http://www.sciencedirect.com/science/article/pii/S1568494612003821>.

[6] Schlechtingen M, Santos I. Wind turbine condition monitoring based on SCADA data using normal behavior models. Appl Soft Comput 2014;14:447–60. http://dx.doi.org/10.1016/j.asoc.2013.09.016.

[7] Chehouri A, Younes R, Ilinca A, Perron J. Review of performance optimization techniques applied to wind turbines. Appl Energy 2015;142:361–88.

[8] Marvuglia A, Messineo A. Monitoring of wind farms power curves using machine learning techniques. Appl Energy 2012;98:574–83.

[9] Kusiak A, Verma A. A data-mining approach to monitoring wind turbines. IEEE Trans Sustain Energy 2012;3(1):150–7.

[10] Schlechtingen M, Santos IF, Achiche S. Using data-mining approaches for wind turbine power curve monitoring: a comparative study. IEEE Trans Sustain Energy 2013;4(3):671–9.

[11] Godwin JL, Matthews P. Classification and detection of wind turbine pitch faults through SCADA data analysis. IJPHM Spec Issue Wind Turb PHM (Color) 2013:90.

[12] Barthelmie R, Hansen K, Pryor S. Meteorological controls on wind turbine wakes. Proc IEEE 2013;101(4):1010–9. http://dx.doi.org/10.1109/jproc.2012.2204029.

[13] Port-Agel F, Wu Y-T, Chen C-H. A numerical study of the effects of wind direction on turbine wakes and power losses in a large wind farm. Energies 2013;6(10):5297–313. http://dx.doi.org/10.3390/en6105297. <http://www.mdpi.com/1996-1073/6/10/5297>.

[14] Hansen K, Barthelmie R, Jensen L, Sommer A. The impact of turbulence intensity and atmospheric stability on power deficits due to wind turbine wakes at horns rev wind farm. Wind Energy 2012;15(1):183–96. http://dx.doi.org/10.1002/we.512.

[15] Gaumond M, Réthoré P-E, Ott S, Peña A, Bechmann A, Hansen KS. Evaluation of the wind direction uncertainty and its impact on wake modeling at the horns rev offshore wind farm. Wind Energy 2014;17(8):1169–78.

[16] McKay P, Carriveau R, Ting DS-K. Wake impacts on downstream wind turbine performance and yaw alignment. Wind Energy 2013;16:221–34. http://dx.doi.org/10.1002/we.544.

[17] Castellani F, Astolfi D, Terzi L. Mathematical methods for performance evaluation of onshore wind farms through scada data mining. Recent advances in energy, environment financial planning. In: Proceedings of the 5th international conference on development, energy, environment, economics (DEEE '14); 2014. ISBN: 978-960-474-400-8.

[18] Castellani F, Astolfi D, Garinei A, Proietti S, Sdringola P, Terzi L, et al. How wind turbines alignment to wind direction affects efficiency? A case study through SCADA data mining. Energy Proc 2015;75:697–703.

[19] Castellani F, Astolfi D, Terzi L, Hansen K, Rodrigo J. Analysing wind farm efficiency on complex terrains. Journal of Physics: Conference Series, vol. 524. IOP Publishing; 2014. p. 012142.

[20] Wagenaar J, Machielse L, Schepers J. Controlling wind in ECN's scaled wind farm. In: EWEA session: innovative wind turbine control; 2012.

[21] Fleming P, Lee S, Churchfield M, Scholbrock A, Michalakes J, Johnson K, et al. The SOWFA super-controller: a high-fidelity tool for evaluating wind plant control approaches. Tech rep. NREL/CP-5000-57175, National Renewable Energy Laboratory, Golden, CO, to be presented at EWEA 2013 Vienna, Austria February 4–7, 2013 (January 2013). <http://www.nrel.gov/docs/fy13osti/57175.pdf>.

[22] Vermeer L, Sørensen JN, Crespo A. Wind turbine wake aerodynamics. Prog Aerosp Sci 2003;39(6):467–510.

[23] Astolfi D, Castellani F, Garinei A, Terzi L. Data mining techniques for performance analysis of onshore wind farms. Appl Energy 2015;148:220–33.

[24] Carln I. Some simple concepts for dynamic wind sector management of wind farms. Tech rep. TG-R-10-06, Teknikgruppen AB; 2010.

[25] Castellani F, Astolfi D, Piccioni E, Terzi L. Numerical and experimental methods for wake flow analysis in complex terrain. Journal of Physics: Conference Series, vol. 625. IOP Publishing; 2015. p. 012042.

[26] Rodrigo J, Gancarski P, Arroyo R, Moriarty P, Chuchfield M, Naughton J, Hansen K, Machefaux E, Koblitz T, Maguire E, et al. IEA-Task 31 WAKEBENCH: towards a protocol for wind farm flow model evaluation. Part 1: Flow-over-terrain models. Journal of Physics: Conference Series, vol. 524. IOP Publishing; 2014. p. 012105.

[27] Moreno P, Gravdahl AR, Romero M. Wind flow over complex terrain: application of linear and CFD models. In: European wind energy conference and exhibition; 2003. p. 16–19.
[28] Porté-Agel F, Wu Y-T, Lu H, Conzemius RJ. Large-eddy simulation of atmospheric boundary layer flow through wind turbines and wind farms. J Wind Eng Ind Aerodynam 2011;99(4):154–68.
[29] Castellani F, Vignaroli A. An application of the actuator disc model for wind turbine wakes calculations. Appl Energy 2013;101:432–40.
[30] Zhong H, Du P, Tang F, Wang L. Lagrangian dynamic large-eddy simulation of wind turbine near wakes combined with an actuator line method. Appl Energy 2015;144:224–33.