A Data-driven Approach for Wind Turbine Performance Bench-marking

Anoop Verma, and Kai Yang

Department of Industrial and Systems Engineering, Wayne State University, Detroit, MI
Email: anoop.verma@wayne.edu

Abstract

Wind turbines consist of several components, sub-assemblies which are prone to fail. Failures in the wind turbines can be broadly categorized into: (1) soft failure, i.e. sensor malfunctioning and (2) hard failure, i.e. actual component failure. Large scale wind farms may contain over hundreds of such fault prone wind turbines. With limited resources, prioritizing the maintenance operations based on the severity and extent of failure is the key to ensure minimum power loss. The research developed here presents a data driven approach for benchmarking performance of utility scale wind farm. Performance of wind turbines is assessed based on turbine power output, and input wind speed. The statistical models are extracted from high frequency operational data which is further utilized to identify performance benchmarks for whole wind farm. Numerous fault and normal operational scenarios are assessed to ensure the applicability of the developed models. The data-driven models developed herein can ensure a continuous performance monitoring to be used for turbine fault prognosis, and maintenance management.

Keywords
Power curve, data mining, statistical moments, control charts.

1. Introduction
Over the past few years, a tremendous growth of wind industry has been witnessed. The growth in the wind turbine has been witnessed in the form of (1) increasing turbine capacity, (2) increasing number of wind farms, and (3) increasing size of the rotor, tower height etc. The performance monitoring and maintenance tasks are thus becoming more crucial than ever. Furthermore, the 20-20 vision of current administration is putting pressure on wind industry to minimize its cost of energy (COE). As per the COE of low wind speed turbines (LWST), O&M costs can account for more than 10% of the total cost [18]. The O&M related costs are expected to grow with the years of turbine operations. A great deal of reduction in COE of wind turbine is thus possible by: (1) developing fault prediction, cost minimization models [18], [16], [8], [17], [9], [11], [20], [6] [7], and [5] (2) increasing turbine power output through efficient/optimum control [15], [13], [14]. The research reported in this paper falls under category (1), where, the COE is minimized by developing an efficient performance monitoring scheme.

The advanced Supervisory Control and Data Acquisition (SCADA) system installed at each turbine typically records hundreds of the parameters that may indicate operational issues in wind turbine. SCADA systems itself able to resolve several operations issues by automatically starting, stopping, and resetting the turbines in case of small fluctuations http://www.renewableenergyworld.com/rea/news/article/2013/ However, further useful knowledge about turbine operations can be extracted by analyzing high frequency data. In a typical utility scale wind farm such data instances grow exponentially, making them almost impossible to analyze altogether. An efficient approach utilizing such high frequency data for near-real/real time assessment is lacking and requires immediate attention. The research presented in this paper addresses such critical aspects of high frequency wind turbine data analysis that can lead to minimization of overall operations and maintenance costs. More specifically, we have developed data-driven wind turbine benchmarking models for performance monitoring.

The paper is organized as follows. Section 2 presents the data set used in the research, along with a brief description of turbine power curves. In Section 3, developed approach is presented. Different performance patterns obtained through developed approach is presented in Section 4. The paper is concluded in section 5 with future research scope.
2. Literature Survey
Wind turbine includes hundreds of components, sub-assemblies which are prone to fail. An efficient way is to measure the impact of such failure is through turbine operations. The operational characteristics of turbines depend on parameters such as rotor power, torque, and pitch angle. Continuous monitoring of these parameters can be useful in assessing wind turbine performance. Candidates for power curve based benchmarking may include: (1) manufacturer’s power curve, (2) physics based models, and (3) data analysis of actual power curve. Due to the aging of wind turbines, component replacement etc., manufacturer’s supplied power curve is hardly being achieved by wind turbines and therefore they are not suitable for benchmarking purpose. Physics based models on the other hand has widely been used. Such models defines the non-linear relationship between wind speed, rotor area, air density, and power output. More specifically the power conversion equation is represented as [1][2]

\[ P = 0.5 \times \rho A \times v^3 \times E \]  

\[ E = E_{\text{rotor}} \times E_{\text{gear}} \times E_{\text{gen}} \times E_{\text{power on}} \]  

Utilizing physics based models, Ref.[10] developed a control chart approach to monitor the progress of wind turbines. In their research, they compared parametric power curve model with non-parametric data driven models. Ref.[19] identified stages of wind turbine power curves for monitoring purposes by transforming the underlying data. They then used the resultant transformed data to detect change detection in turbine performance. Ref. [1] developed a heuristic for setting alarm limits in power curve. They used their model for rotor condition monitoring. Ref. [2] on the other hand used performance indicators namely availability, windiness, long-term wind speed and power performance in order to evaluate the overall wind farm performance. The above mentioned literature clearly indicate the role of power curve based analysis in turbine performance monitoring. However, most of the analysis in targeted to an individual wind turbine. Due to individual differences in the turbine performance, models built on one turbine may not be applicable to another. A generalized framework for power curve based wind turbine performance monitoring is required. The research developed here in an extension of authors’ previous work [4].

In next section, details of developed solution scheme is presented.

3. Developed Approach
We have developed a holistic approach to identify the reference curves for benchmarking. The approach starts with appropriate data selection, which is later on smoothed for referencing. The high-frequency dataset is then transformed into meaningful metric which forms the basis for performance comparison (See Figure 1). The results of the analysis are then validated with turbine operational data containing information about the production output, availability etc. The salient aspects of the developed framework is discussed next.

![Diagram of the overall approach for continuous performance monitoring](image)

Figure 1: The overall approach for continuous performance monitoring

3.1 Dataset description
The dataset used in present research is obtained from a utility scale wind farm, consisting of over 100 wind turbines (1.5 MW capacity). The power curve information of 22 representative wind turbines is used in the analysis. A four
year historical power curve data (Aug 05- Aug. 09) of 22 wind turbines was available for the analysis. The overall dataset is divided into three parts for reference curve construction, model testing, and validation respectively. Table 1 display the distribution of the available data. The selection of wind turbines for the analysis was based on the data completeness. As the performance of a wind turbine largely relies on the input wind speed, dataset depicting the generalized model of the wind turbine should be chosen with care. A unified reference curve for whole wind farm is ideal, however, due to the variability in the wind speed across each month, reference curve for individual months were extracted.

<table>
<thead>
<tr>
<th>Sl. No</th>
<th>Task</th>
<th>Data</th>
<th>No. turbines</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Reference curve construction</td>
<td>Aug. 05-08</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>Reference curve testing and validation</td>
<td>Aug. 09</td>
<td>01</td>
</tr>
<tr>
<td>3</td>
<td>Continuous monitoring</td>
<td>Jan-Dec 09</td>
<td>10</td>
</tr>
</tbody>
</table>

### 3.2 Wind turbine power curve

Power curve is unarguably the most significant and informative aspect of a wind turbine. It describe the relationship between the input wind speed and the power output and usually form sigmoid shape for whole wind speed range. Abnormalities, machine maintenance, faults makes the power curve of wind turbines to differ from each other. A typical wind power curve in general resembles to a sigmoid function, however due to various malfunctions, e.g., sensors and components, the power curve acquires its own shape. The power curve supplied by the turbine OEM is typically ideal, however, due to performance degradation over time, same can not be used for the analysis. In the present analysis, the high frequency 10 sec dataset is averaged over 10 min across 22 wind turbines for better generalization. Figure 2 display the comparison between a single turbine and 10 min average of 22 wind turbines. The under-performing data points are averaged out, providing a better power curve for reference. The power curve however still contains abnormal data points (data points with no power at high wind speed) and needs to be removed. In the present research, an approach based on Mahalanobis distance is used to remove such abnormal data points, discussed in section 3.2.1.

#### 3.2.1 Outliers detection

Here the term outliers refers to the instance where the turbine operations was abnormal, including under-performance. Such instances include: (1) high power at low wind speed, low power at high wind speed, and (2) relatively low power at given speed. Removing such instances is useful while extracting a reference curve from the averaged power curve. Due to the nature of the curve, outliers in the power curve are extracted and hence removed using Mahalanobis distance information. Mahalanobis metric expresses the distance of an instance to the centroid in the multidimensional space.
D_2^{st} = (x_s - x_t) \text{cov}^{-1} 1(x_s - x_t)^{-1}, s \neq t \tag{3}

In equation (3), $D_2^{st}$ is the Mahalanobis distance between instance $s$, and $t$, and $\text{Cov}^{-1}$ is the inverse of covariance matrix. Figure 2 display the outliers of the power curve given in Figure 2 (b). A k-means clustering is initially applied to get outliers for each zone (i.e. low wind speed, mid wind speed, high wind speed etc.). A hard thresholding is applied on the partial power curves to extract a smooth curve.

4. Turbine performance metric

The large instances of dataset makes it computationally challenging for the analysis. Data compression based on first and second order moments reduces the instances, however, at the expense of information loss. In the present research, third and forth order moments were selected as the turbine power curves typically form a distinct curve. The third and fourth order moments namely kurtosis and skewness are often used to describe the shape of the data distribution. The literature pertaining to usage of skewness and kurtosis metric is primarily related with depicting normality in multivariate analysis. Other applications of kurtosis and skewness includes (1) identifying initial component in independent component analysis (ICA), and (2) identification of dynamics in N-dimensional market. The refined reference curves obtained in the previous section are used as a benchmark for continuous performance monitoring of the wind farm. The multivariate skewness, kurtosis is defined in Equations (4) and (5) [12], [3].

$$S_{bi} = \frac{1}{n^2} \sum_i \sum_j ((x_i - \bar{x})^T \text{Cov}^{-1} (x_j - \bar{x}))^3 \tag{4}$$

$$K_{bi} = \frac{1}{n^2} \sum_i \sum_j ((x_i - \bar{x})^T \text{Cov}^{-1} (x_j - \bar{x}))^2 \tag{5}$$

4.1 Hypothesis testing

In general, under-performing wind turbines will deviate from the reference curves, resulting in different values of kurtosis and skewness values. Figure 4 display the comparison of monthly and yearly skewness-kurtosis metric. A monthly testing instance is selected for comparison. It is clearly depicted from figure that the metrics for monthly reference curves are more closer to monthly power curve than yearly reference curve and therefore ideal for performance benchmarking. The power curve of twenty-two wind turbines for same month are analyzed and their transformed matrices are calculated, shown in Figure 5. The symbol $\star$ represents the reference point, whereas, $\diamond$ represent the individual turbine status. Figure 5 also displays the power curve of distinctly identified turbines for validation. Turbines located farthest from the reference points are considered to be operating abnormal and would need immediate attention. Depending on the distribution of data points across the power curves, the actual performance point varies and can be tracked: (1) individually for a turbine over a period of time, and (2) for whole wind farm. In 2-D skewness-kurtosis graph, wind turbines performance can be assessed by: (1) relative location of individual turbines with respect to the reference curves, and (2) location of individual turbines with respect to the turbine clusters. In general, turbine
Verma and Yang

Figure 4: Comparison between monthly and yearly reference curve metrics

Figure 5: Monthly performance status of 22 wind turbines

showing the same behavior will form a distinct cluster. Any abnormal turbine behavior can be easily visualized in a 2-D scatter graph. The possible reasons for distinct location of individual turbines in skewness-kurtosis plot could be: (1) under-performance due to system abnormalities, (2) under-performance due to different wind speeds, (3) overperformance due to errors in wind speed measurement. Using the guidelines mentioned earlier in this section, power curve based skewness-kurtosis graph identifies turbine 10, and turbine 12 as abnormal, whereas, turbine 13, turbine 17, and turbine 9 behave differently in the rotor and blade pitch curves. The abnormal behavior of turbine 10 is clearly visible as the fault logs confirm the faults associated with generator windings. The fault log data recorded by the SCADA systems confirm the faults associated with generator windings. More information about turbine fault logs is provided in [16].

In next section, continuous monitoring scheme developed from transformed performance metrics is discussed.

5. Continuous monitoring

The metrics obtained from reference curves typically represents an ideal scenario of turbine performance for the corresponding month. An xbar control limit of the transformed reference curves metrics is developed to signify the performance tolerance zone of the wind turbines. The upper and lower control limit of both bivariate skewness and kurtosis is obtained following an xbar chart. Figure 6 display the control limits of both skewness and kurtosis metrics of the testing dataset. The actual performance metrics found to lie well within the control limits.
6. Performance patterns

Based on the identified control limits for individual months, the power curve of 10 wind turbines over a period of one year (Jan 2006-Dec 2009) is analyzed. Figure 6. The trend is visually observed for each month and three distinct patterns were found (see Figure 7). In pattern 1, all wind turbines were found to be well within the limits, thereby declared as normal scenario, whereas, in case of pattern 2, all but 1 turbine (i.e. turbine 2) was found to be well outside the limit. In that case, turbine 2 was found to have some malfunctions. In case of performance pattern 3, all wind turbines was found to be outside of the identified limit. A typical reason for such case could be due to the external factors such as shutdown due to extreme weather, or curtailment. In order to get a better insight of the performance patterns, the monthly production data of the corresponding months was analyzed. The production data contains the information about the overall monthly production of wind turbines, its availability, and time when the turbine was actually operational (see Figure 8). The capacity factor (CF) and effective capacity (EC) of the wind turbines were calculated. Capacity factor is a well known metric in wind turbine industry that indicates the amount of energy a wind
Figure 8: Sample production data of 10 turbines.

Verma and Yang

turbine produces, whereas, the effective capacity metric is developed by authors which represent the amount of energy a particular wind turbine produces when it is operational. The EC metric is developed to figure out the root cause of performance degradation as shown in Figure 7 (c). Mathematically, the CF and EC can be written as (see equation 6 and 7):

\[
CF = \frac{\text{Production (KWh)}}{\text{No. of days} \times \text{no. of hours/day} \times \text{turbineratedcapacity}}
\]

(6)

\[
EC = \frac{\text{Production (KWh)}}{\text{actualoperationaltime} \times \text{turbineratedcapacity}}
\]

(7)

Figure 9 display the plot of CF and EC of the monthly performance patterns of 10 wind turbines, shown in Figure 7 (a-c). In case of Figure 9, both EC and CF was found be very close, thus confirming normal operation. In case of Figure 10, the CF of turbine 2 was found be lower than EC and other turbines. This indicates turbine 2 was under-performing. In case of Figure 11, however, there was a significant difference in the CF and EC, however, the trend was found to be similar across each turbine. This case typically represent an external effect. The further investigation of fault logs confirm the curtailment of wind turbines as the demand of the particular month was already met.

Figure 9: Comparison between CF and EC of wind turbines of month May 09

7. Conclusion and Future Research

A holistic approach for monitoring performance of a wind farm was presented. Power curve was used as a reference curve for performance comparison. The results developed in the present paper were promising as the high frequency data was transformed into low frequency (a singular value) without loss of much information. The models developed in the present paper was tested and validated on monthly dataset. Extending the applicability of present approach to weekly or daily data is a topic of future research. Linking the fault instance with the performance patterns will also be explored in the future.
Verma and Yang

Figure 10: Comparison between CF and EC of wind turbines of month Jan 09

Figure 11: Comparison between CF and EC of wind turbines of month Sep 09

References


Verma and Yang


