Unsupervised Learning and Fusion for Failure Detection in Wind Turbines

Xiang Ye, Kalyan Veeramachaneni, Yanjun Yan and Lisa Ann Osadciw
Department of Electrical Engineering and Computer Science
Syracuse University, Syracuse, NY 13244-1240
(315)443-3366 (office)/(315)-443-2583 (fax)
xye/kveerama/yayan/laosadci@syr.edu

Abstract - Unforeseen failures of components of a wind turbine have a significant impact on the power generation capability. Anemometer, which measures the wind speed as seen by the turbine, can be used to predict within some degree of accuracy the amount of power the turbine should produce. It is a good indicator of incipient failures that the power production is consistently low. Unsupervised and automated monitoring of the dynamic behavior of a wind turbine and power produced can detect faults early. Thus, secondary defects and major breakdowns are avoided as well as minimizing the impact this failure has on the total production of the turbine. This paper presents a failure detection method relying on the wind speed differences between two turbines. We collect a week’s worth of data and model the wind speed difference between two turbines using a Weibull distribution. Two original tests are proposed. The first test is based on the fitted Weibull distribution, where the abnormal weeks have higher values in both the scale and shape parameters. The second test evaluates the area under the weekly empirical cumulative density function. Both tests yield similar fault detection results, and the second test reveals the sequential progression. A fusion using the second test on multiple turbine pairs can clarify the ambiguity in the single turbine pair test.

Keywords: Failure detection, data fusion, Weibull distribution, multivariate analysis, condition monitoring

1 Introduction

Wind energy has been receiving more attention in the last decade. It will play an important role in the future energy market since the decreasing production of the fossil energy. However, wind energy cannot compete with the traditional energies up until now since it is more expensive than others. Even though the availability can be up to 98% for wind turbines, the cost for maintenance is still very high. For the wind power systems, the availability may fall below 60% due to the downtime of the wind turbines, which is mainly caused by the failures of some components of wind turbines [1]. Therefore, to make wind energy more competitive in the future, efforts are required to enhance the availability, reliability and lifetime of the wind turbines.

One of the approaches to reduce the cost for maintenance is the early failure detection. If failures are detected at an early stage, the consequent damage is minimized or mitigated, and also repairs are better scheduled. This leads to shorter downtimes and lesser revenue losses.

Many common failures in wind turbines such as the bearing and gearbox failures are studied thoroughly by many manufacturers, and their diagnostic units are built into the wind turbines. These are hard failures since they cause the turbines to totally shut down. However, many soft failures, which degrade turbines’ performance but still keep them running, such as anemometer faults are overlooked, which are also very harmful to the efficiency of the turbines. Anemometer measures the wind speed as seen by the turbine. The measurement enables the configuration and control settings of the turbine. Turbines are usually shut down for a wind speed above 20 m/s for safety reasons. Also, the maximum power production is usually achieved, which is approximately 15-18 m/s. A faulty anemometer reading will shut down the turbine when it can actually produce energy. Similar, a faulty reading can also cause severe damage if turbine is not shut down at right time. Further, wind speed is correlated with other measured variables such as rotor speed, power produced, etc., for online monitoring of the turbine. A faulty wind speed reading will cause unnecessary false alarms and frequent shut downs causing severe downtime. Hence for successful operation of the wind turbines it is paramount that we monitor anemometer and detect faults in it.

In our detector design, we choose two neighboring turbines in their physical locations, whose ambient wind speed should be similar. If the measured wind speed is drastically different, it could be due to anemometer fault or due to different turbulence at different locations, noisy SCADA data or other factors such as wind blockage. A direct comparison of point-to-point wind speed differences easily triggers false alarms. Therefore, a week worth of wind speed data is aggregated for robust detection. The measured wind speed differences in a week are estimated by a two-parameter Weibull distribution, which readily reveals the normal and abnormal region of the parameters. With the 2D plot of Weibull scale and shape parameters, both hard failures and soft failures are obviously detected.

Further, we pragmatically derive the cumulative probability distribution function (CDF) for each turbine pair. Suppose that the absolute wind speed difference is \( x \), and the
CDF is $F(x)$. We choose a common value of wind speed difference, $d=3$ meters/second, to evaluate the CDF of each turbine pair at $d$. $F(d)$ serves as an indicator on how many wind speed difference data points fall below $d$ in each week. The more the data points fall below $d$, the less their difference is, and hence the less likely there are faults. The less the data points fall below $d$, the bigger their difference is, and the more likely there are faults. $F(d)$ is a scalar, and this detector simplifies the decision and enhances the sensitivity of fault detection.

By utilizing the general SCADA measurements, we propose a data-driven approach to identify under performance events, some of which are consistent with the flagging of the manufacture diagnostic units; yet others are not flagged. Further study of the unflagged data shows that there are still failures, including the anemometer problem. We find that the measured wind speed difference between neighboring turbines is critical for such detection, and it is used in our detector.

The rest of the paper is organized as follows. In Section 2, we describe the wind energy system problem. Then, we explain the failure detectors between two turbines in Section 3. We evaluate the technique’s performance among multiple turbines in Section 4. Finally, concluding remarks are drawn in Section 5.

2 Problem Description

There are several wind turbines located on the site with obstructions and local topographic features, which would introduce new factors affecting the performance measurements. The performance data is collected by the SCADA system with different types of sensors from each turbine, and the data is recorded every unit time interval.

The data currently available from SCADA includes not only measurements, but also some flagging of events or faults from the existing built-in diagnostic units. However, such flagging is designed for some failures, but it is not comprehensive in detecting other faults, especially the degradation of the faulty components. If the abnormal behaviors of the turbines are caught early, an early fix can not only improve the turbine efficiency but also save on costly repairs for a major fault later. Hence, we propose to use a data-driven approach to take advantage of the available measurements without installing extra sensors explicitly for diagnostics to see how we do. The usage of the general performance data in a sophisticated detector design is shown to be able to detect abnormal behaviors effectively. The advantages of using general performance monitoring data for diagnostics is that it is more cost effective and can be applied to virtually any turbine.

3 Failure Detector Between Two Turbines

An anemometer measures the wind speed as seen by the turbine. This measurement is affected by several factors like elevation, turbulence at the turbine, other topographical factors or simply noise. However, it is fair to assume that two neighboring turbines would most likely measure a similar wind speed with a slight offset $\delta$. A difference significantly greater than this offset is an indicator of an anemometer failure, turbine malfunctioning or wind blockage. A deterministic model for this offset is hard to derive. Also, in absence of a model for faulty functioning of an anemometer, we cannot devise a supervised scheme to detect the failure. In this paper, we rely on probability density function derived for a week worth of data and track the shifts of the data using this model. If the probability density function for the wind speed difference has a large deviation from our model, it is indicative of a failure.

3.1 Probability Distribution of Wind Speed Difference Fitted with Weibull Function

We choose the turbine A and its closest neighboring turbine B and evaluate the difference in measurement of wind speed between these two turbines. We divide the whole data set into weekly data, with “n” samples per week. A week worth data is enough to gain accurate estimation, and not too much to smooth out the failure features. For each week, we find the absolute values of wind speed difference between A and B sample by sample. This variable is given by

$$x = |ws_A - ws_B|$$  \hspace{1cm} (1)

Since Wind speed as seen by the turbine follows a Weibull distribution [3], we estimate the weekly data of wind speed difference between the two turbines to a Weibull distribution given by

$$f(x; \lambda, k) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}$$  \hspace{1cm} (2)

where $x$ is the Weibull random variable given by (1), and both $k$ and $\lambda$ are greater than zero. The shape parameter $k$ and the scale parameter $\lambda$ for the Weibull distribution are estimated for each week’s data.

For example, Figure 1 and Figure 2 show the wind speed difference histogram and its estimated Weibull density function for two different weeks, i.e., week 31 and week 79, respectively. Since ideally we would like both turbines to see identical wind speeds, we can say the lower the spread of the Weibull distribution, the better the functioning of the anemometer in the two turbines. We can see from Figure 1 and Figure 2 that there is a significant difference between the spread of the two distribution functions.
3.2 The Good, The Bad and The Ugly Patterns of Wind Speed Difference

Figures 1, 2 and 3 demonstrate the “good”, the “bad” and the “ugly” Weibull wind speed difference density function, respectively. We consider a Weibull density with a shape and scale parameter as a single state for the wind speed difference variable under analysis. Figure 1 is the state for the week 31, figure 2 is the state for week 79 and the figure 3 is the state for week 118. The one in week 31 has a sharper shape distribution. The value of the wind speed difference is less than 2 m/s, and the probability of zero wind speed difference is close to 1. So, it has a better performance since we have a good estimate of the wind speed difference with a high probability. If we need to predict the wind speed difference within 1 m/s, we know the difference will be 1 m/s with nearly 100% probability. However, the distribution for week 79, shown in the Figure 2, is spread widely between 0 m/s and 6 m/s. We could not select a wind speed difference within 1 m/s that week. This distribution is indicative of a fault on one of the two anemometers. However, we cannot conclusively say, which turbine in fact has an anemometer problem without a second wind speed difference test.

For week 79, the shape parameter, $k$ is greater than 1. But for week 118, this is lower than 1. This implies that the shape of the distribution is quite different for the two weeks. Figure 3 shows the shape of the distribution for week 118. We can see that although the distribution is spread out but the probabilities for wind speed difference higher than 4 are significantly lower. When we inspected the maintenance logs we found a turbine shut down due to lightning or a spindle failure. An intuitive explanation for the shape of the distribution can be that if one of the turbines measurements is zero, then we would see a distribution pertaining to a single wind speed, which is spread out with low probabilities for high wind speeds. For week 79, by inspecting the maintenance logs we found certain failures which are detailed later in the paper.

As discussed above, the probability distribution of weekly wind speed differences is determined by its estimated Weibull parameters, the shape parameter $k$ and scale parameter $\lambda$. So, we can detect turbine failures based on the pairs of Weibull parameters, which also determine the good and bad states of wind speed difference.

3.3 Two-Dimension Failure Detector by Plotting Estimated Weibull Parameters of Weekly Wind Speed Difference

We analyze the probability density functions for the 130 weeks worth of data. The Figure 4 shows the 2-dimension plot of the estimated Weibull scale and shape parameters of weekly wind speed difference between the turbine A and B. The red spot is the cluster center of all parameters. A circle with this center represents the normal functioning of the anemometers on both the turbines. In fact, the circle can act as a detection threshold for “bad” and “ugly” states. The weeks inside the circle pertain to good states, when both turbines are functioning properly, and the ones outside the circle are functioning less optimally in a “bad” or “ugly” state. Notice that week 79 and week 118 are both outside the circle. The two anomalous points in 2-D space are indicative of faulty functioning, however, with two distinct
3.4 Understanding System Failure by Integrated Wind Speed Difference Cumulative Distribution Function

To improve our determination of system failure, we propose another test for the wind difference, which fuses these two parameters. Given the estimated Weibull parameters of the wind speed difference for each week, we evaluate the Weibull cumulative distribution of the wind speed difference as

\[
F(x; \lambda, k) = 1 - e^{-\left(\frac{x}{\lambda}\right)^{k}}
\]

where \(x\) is the Weibull random variable of wind speed difference, and \(k\) is the shape parameter, and \(\lambda\) is the scale parameter. If both two turbines working well, the cumulative probability of the wind speed difference between them should rise to a value 1 faster. In such a case, the two turbines rarely have different wind speeds. Otherwise, if one turbine is affected by failures, the Weibull cumulative distribution approaches a diagonal or nearly linear function, which corresponds to a uniform distribution. For example, as shown in Figure 5, the Weibull cumulative distribution of wind speed difference in week 31 is much steeper than those in week 79 and week 118.
When the cumulative density function of the Weibull distribution rises sharply, the curve covers much more area under it. Also, from figure 5 we notice that the CDF for week 118 rises faster than the CDF for week 79. Based on this observation, we consider the area under Weibull cumulative density function of the wind speed difference for each week as a good parameter to understand the validity of the wind speed measure for that turbine. In Figure 5, the area of week 79 has the lowest value. Figure 6 plots the area in the histogram of the real CDF as a function of week and plots this area under the CDF curve as a function of the week numbers. A failure in the anemometer or any turbine sub-system involved in the wind measurement would cause the same change in the turbine subsystem.

In our first detector discussed in section C, we considered a threshold quadrant defined with the shape and scale parameters bigger than 0.9, This set of parameters is translated to a CDF area of 114, which is used as the threshold in Figure 6, the green straight line, to declare abnormal weeks, and the same detection result as by the first detector is obtained.

### 4 Failure detector by fusing information from multiple neighboring turbines

Once a failure is detected using the previous methods, it cannot be conclusively said which of the two turbines have an abnormal condition. To further enhance the performance of our detector we evaluate the wind speed differences with respect to multiple turbines for the turbine under test. We consider the turbine A and evaluate its wind speed difference with respect to B, C, D, E, F, G and H. Order these turbines numerically with a notation of $i$. For instance, turbine B is the 1st turbine in comparison, turbine C is the 2nd turbine in comparison, etc. until turbine H is the 7th turbine in comparison. In our example, we compared the wind speed of turbine A with seven other turbines. In practice, we can use more or less turbines for comparison. For $i$ goes from 1 to 7, the $i$th wind speed difference is given by

\[
    x_i = |w_{x_A} - w_{x_i}| \quad (4)
\]

Similar to our analysis in pervious sections we estimate the probability density function and the cumulative density function for each of the wind speed difference. Further we will attempt to fuse these multiple sources of information to determine whether turbine A has an abnormal condition or not. If turbine A functions abnormally, then it is highly likely that some or all of the wind speed difference CDFs are in bad states. On the other hand, if turbine A functions properly, then it is unlikely that all of the CDF tests would be bad.

Due to the differences in the physical distances, wakes, or potential faults, the wind speed difference between multiple turbine pairs would differ. The CDF plots of the weekly wind speed difference is averaged over the full data duration, and these average CDF plots between turbine A and multiple turbines are shown in Figure 7 to show the variation. Note that the best in this average is turbine B. The bigger the area under the CDF plot, the smaller the difference between the wind speeds measured on the turbine pair. Hence, it is the more important this turbine in comparison in telling when turbine A may function abnormally. Based on this idea, the areas under the average CDF curve are used as the weights for later fusion.

\[
    w_i = \int_0^\infty CDF(x_i)dx_i
\]

where $x_i$ is defined in Equation (4).

For each turbine pair, the areas under the empirical weekly CDF curve are first evaluated. In the $j$th week, the $i$th turbine assumes an area under the CDF curve as below.

\[
    A_{i,j} = \int_0^\infty CDF(x_{i,j})dx_{i,j}
\]

(6)
where $x_{i,j}$ is the wind speed difference between the turbine A and the $i$th turbine in the $j$th week. Then the maximum area within all these weeks is used to normalize the weekly area, so that the normalized area is in the range of 0 to 1.

$$a_{i,j} = \frac{A_{i,j}}{\max_j(A_{i,j})}$$

where $\max_j(A_{i,j})$ takes the maximum value over week number $j$.

Finally the multiple turbine tests are fused as below.

$$FA_j = \sum_{i=1}^{m} w_i a_{i,j}$$

where $m$ is the number of turbines in comparison. In our example, $m=7$. $FA_j$ is the weekly fused result, as shown in Figure 8.

In figure 8, the test result between the single turbine pair, in red line with circle markers, is taken from Figure 6, which is compared with the fused result, in blue line with star markers, between the multiple turbine pairs.

From Figure 8, we can see that the single turbine pair test indicates that week 79 may have a problem. With the multiple turbine pairs test, this week is excluded. It indicates that the low value in the single turbine pair test in week 79 is due to turbine B, instead of turbine A. With a single turbine pair test, we cannot conclusively say which turbine causes the problem. However, with the multiple turbine pairs test, we can determine which turbine is faulty in that week.

5 Conclusion

This paper proposes two failure detectors based on wind speed difference between a turbine pair. Wind speed difference is a good indicator of failures related to anemometers directly or indirectly.

The wind speed difference conforms to Weibull distribution, and the Weibull distribution parameters are estimated from weekly data. The first detector uses the parameter values to determine the normal/abnormal region. The detected abnormal weeks are found to be faulty.

The second detector calculates the area under the empirical cumulative distribution function. If the two turbines are consistent, which means that they both perform properly, then their wind speed difference is small, and the area under the CDF is large. A single pair turbine test picks out the abnormal weeks, which can be caused by either the first turbine or the second turbine. With a turbine of interest as the reference, a multiple turbine pairs test can further exclude the ambiguous detection in the single turbine pair test. The second detector shows the sequential trend, which is also helpful for early detection.

Acknowledgements

AWS Truewind for their generous support financially and technically to this research.

References

[1] M. Verhaegen, “Model Based Fault Detection for Large Scale Wind Turbines”.


