Predicting Wind Turbine Blade Erosion using Machine Learning

Casey Martinez  
*Southern Methodist University*, casianom@mail.smu.edu

Festus Asare Yeboah  
*Southern Methodist University*, fasareyeboah@smu.edu

Scott Herford  
*Southern Methodist University*, jherford@mail.smu.edu

Matt Brzezinski  
matt.brzezinski@patternenergy.com

Viswanath Puttagunta  
vish@divergence.ai

Follow this and additional works at: [https://scholar.smu.edu/datasciencereview](https://scholar.smu.edu/datasciencereview)

Part of the [Applied Statistics Commons](https://scholar.smu.edu/datasciencereview), [Engineering Commons](https://scholar.smu.edu/datasciencereview), and the [Longitudinal Data Analysis and Time Series Commons](https://scholar.smu.edu/datasciencereview)

Recommended Citation

Available at: [https://scholar.smu.edu/datasciencereview/vol2/iss2/17](https://scholar.smu.edu/datasciencereview/vol2/iss2/17)

This Article is brought to you for free and open access by SMU Scholar. It has been accepted for inclusion in SMU Data Science Review by an authorized administrator of SMU Scholar. For more information, please visit [http://digitalrepository.smu.edu](http://digitalrepository.smu.edu).
Predicting Wind Turbine Blade Erosion using Machine Learning

Casey R. Martinez, Festus A. Yeboah, James S. Herford, Matt Brzezinski, Viswanath Puttagunta

1Master of Science in Data Science
Southern Methodist University
Dallas, Texas USA
{casianom,fasareyeboah,jherford}@smu.edu,
matt.brzezinski@patternenergy.com, vish@divergence.ai

Abstract. Using time-series data and turbine blade inspection assessments, we present a regression model used to predict remaining turbine blade life in wind turbines. Capturing the kinetic energy of wind requires complex mechanical systems, which require sophisticated maintenance and planning strategies. There are many traditional approaches to monitoring the internal gearbox and generator, but the condition of turbine blades can be difficult to measure and access. Accurate and cost-effective estimates of turbine blade life cycles drive optimal investments in repairs and improve overall performance. These measures drive down costs, as well as, provide cheap and clean electricity for the planet. It has been shown that by sampling actual blade conditions with drone photography, the remaining turbines can be predicted with acceptable accuracy (86% R-Squared Error). This result not only saves money on drone inspections but also allows for a more precise estimation of lost revenue and a clearer understanding of repair investments.

Keywords: Wind Energy, Predictive Maintenance, Blade Erosion, Machine Learning

1 Introduction

As a form of solar energy, wind is caused by three concurrent events: the uneven heating of Earth’s atmosphere by the sun, irregularities found on Earth’s surface, and the rotation of the Earth. Across the United States, the variation of wind flow patterns and speeds are substantial and greatly modified by bodies of water, vegetation, and differences in terrain. Both “wind energy” and “wind power” describe the overall process where wind is leveraged for the purpose of generating mechanical power or electricity; the kinetic energy harnessed from wind turbines is thereby transformed into mechanical resulting in specific tasks, such as grinding grain, pumping water, or using a generator to convert mechanical power into electricity [1].
The production of electricity from energy captured by wind turbines stems from the aerodynamic force created by rotor blades. As wind flows across the wind turbine blades, air pressure on one side of the blade decreases. The difference in air pressure across the two sides of the blade creates both lift and drag [1]. Because the force of the lift is stronger than the drag, the rotor shaft is then forced to spin. The generator contained within the turbine is connected to the rotor shaft.

Modern-day wind turbines are complex and sophisticated machines with hundreds of moving parts. A team of technicians and engineers are needed to maintain operation at levels needed to provide cheap electricity to market. Most of the rotating components inside the nacelle portion of the turbine have temperature and vibration sensors. These sensors alert the control system of anomalous equipment behavior. Since the fiberglass blades themselves have no moving parts and can be 50 meters in length or greater, similar monitoring systems have not been implemented at scale. The industry relies on periodic inspection via telescoping photography, drone photography, and/or human inspection with ropes and harnesses. Each of these approaches require turbine downtime, which can require labor intensive tasks, such as collecting large volumes of data and making blade assessments. An example of traditional blade inspection performed by hand with a technician rappelling down each blade (Figure 2).
Since it is not cost effective to inspect every turbine in the fleet, only a sample of turbines are inspected. These inspections can also be prone to human error. This error is due to the subjective nature of the assessments, which are usually based on photography. The defect of interest in this work is the erosion of blade materials on the leading edge. This failure mode is a slow process which can etch away the gel coating finish of the fiberglass blade. Other defects caused by manufacturing defects, lightning strikes, and other events are not considered here.

There are numerous defect types that can appear on a blade from cracks and pitting to full delamination in sections. The focus of this paper is on leading edge erosion. Most of these defects are either caused by an acute event such as lightning, blade erosion is caused by small physical impacts over time. This defect type can be difficult to see in early stages and can have a significant impact on power generation. Sources of particles in the air can be dust, rain, bugs, and hail. Leading edge blade erosion caused by water droplet impingement reduces blade aerodynamics and limits power output. If left uncorrected, the structural integrity of the blade can be compromised.

2 Background

Leading-edge blade erosion has become a common problem among wind farm operators. Better prediction of blade erosion, so maintenance can be planned and scheduled, can produce a positive financial impact[2]. The entire wind energy sector is faced with effects from the erosion of materials used in wind turbine blades,
which are primarily due to the effect of weather conditions[3]. Blade material erodes due to the effect of rain, hailstones, and sand dust, which significantly reduces the service life of wind turbines. Accelerated replacement of turbines becomes expensive: up to 2 to 4 percent of the value of all wind-generated power is lost as a result of this problem[3].

The blades consist of a fiberglass frame and shell with a gel coating on the surface. These materials are used for their desirable weight to strength ratio. This gel coating can be damaged by debris in the air such as dust, rain, hail, and other materials. This is due to the high tip speed of the blades. For an average wind speed of 13 m/sec, the tip of the blade can move at speeds approximately 120 mph. At this speed, the gel coating can be eroded over time. Figure 3 shows an example of leading edge erosion.

![Fig. 3. Example of Leading Edge Erosion](image)

The blades of the turbine are the most difficult part of the turbine to collect data from. Modern turbines have sophisticated condition monitoring systems for the rotating equipment inside the nacelle. The expensive rotating equipment, such as the generator and gear box, are monitored by temperature and vibration sensors where damage can be flagged quickly. No direct measurements of the blade aerodynamics or structural integrity is collected by the turbine. Turbine damage has to reach a significant level before a loss in aerodynamics, and therefore power efficiency, can be detected. Analytical methods to use this reduction in turbine efficiency will be discussed in a later section.
Blade inspections have seen a rapid advancement due to autonomous drones and burst photography. In some estimates, 85 percent of blades currently installed have defects[4]. Wind farm operators need a way to prioritize and estimate repair costs. As can be seen in figure 4 below, automated blade inspection based on drone photography can reduce human error and lead to more consistent results.

The blade inspection ratings, physical damage description, and corresponding impacts to turbine performance can be see in Table 1. Turbine performance is not affected until a score of three and above. If this change in turbine aerodynamic efficiency can be directly observed in the turbine’s performance, perhaps it can be used to help predict the turbine damage. This under-performance can be estimated using the first year of a turbine’s operational life as a baseline. This baseline assumes pristine condition of the blades and that the turbine has seen a wide range of environmental conditions. Then, if erosion does exist, the gradual drop in performance will be revealed by the divergence of the actual and predicted performance, called residual. This performance residual is a key feature to predicting blade erosion.

Traditionally, turbine performance was analyzed by careful study of the turbine’s power curve. This is the relationship between wind speed and electrical power output. The theoretical power curve is modeled by the turbine supplier using the turbine power generation equation. This simple representation does take air density into account but does not include other factors such as air speed uniformity, known as turbulence, and vertical air flow, known as sheer. The air
density, turbulence, and sheer can be calculated using the meteorological tower data installed at the wind farm[5].

Factors such as terrain, wind direction, wind shear, turbulence, air density and waking effects of other turbines can cause a significant amount of noise in the power curve. Accurate modeling of turbine power performance cannot be properly captured in this two-dimensional power curve of wind speed and power generation. A multivariate regression modeling approach is used.

3 Literature Review

Over the last few decades, various methods have been employed to predict the performance and condition of wind turbine monitoring systems; however, the majority of faults continue to be detected through planned maintenance sessions [6]. Several studies have used Artificial Neural Networks (ANN) to model the normal behavior of wind turbines; thus, ANNs are being used to monitor performance in real time and predict wind power performance, which can be leveraged for the purpose of optimizing fault detection systems [6]. Costa et al. provide several models that have been used extensively for short-term prediction [7]. For example, they present use cases of multiple time series and neural network methods in order to output energy demand as well as energy product estimates. According to Costa et al., these methods can be purposed for bringing significant operational advantages to energy production grids like wind turbine farms [7].

As mentioned above, the SCADA system has been utilized for collecting and storing data from wind farm operations. In order to manage power consumption and turbine efficiency, the SCADA system was integrated with turbine control systems to detect downtimes [8]. Further, data from SCADA has been used for classifying turbine failure events into categories separated by level of severity using statistical methods to determine reliability of wind farms [9]. D. Fawzy et al. leveraged the Wind Turbine Erosion Predictor (WTEP) to predict erosion rates of wind turbines while accounting for environmental factors [10].

4 Data Formatting

The data collected for this paper consists of three types: blade inspection data, turbine performance data, and meteorological data.

4.1 Blade Inspection Data

Blade inspection data was collected via drone photography by two contracting companies. Each blade was imaged on all sides after the blade was stopped and moved to the vertical position. These images were later processed and automatically classified using an image classification AI technique. We believe the classification uses contrasting pixels. All scores above three are reviewed by a human to check for proper scoring and to provide additional commentary.
The altitude of the drone was collected during the inspection. Since the blade is locked in the vertical position during inspection, the altitude corresponds to the position along the length of the blade. The location of the defect along the blade length is important to target different types of blade defects.

Table 1 shows how defects are scored and if turbine performance is impacted due to changes in the aerodynamic efficiency of the blade. As you can see, when the gel coat and fiberglass of the blade is damaged, the performance of the blade is reduced. Level five damage can cause blades to crack and break if not repaired. Such broken blades can strike the tower and cause the entire turbine to fall to the ground.

Table 1. Overview of Inspection Scores

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
<th>Performance Impact?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cosmetic Damage</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>Some Spotting</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>Gel Coat Damage</td>
<td>Yes</td>
</tr>
<tr>
<td>4</td>
<td>Fiberglass Damage</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>Structural Damage</td>
<td>Yes</td>
</tr>
</tbody>
</table>

According to the pie graph in figure 5, 85% of the defect data consists of level one and two defects.

Fig. 5. Distribution of Severity Classes
The inspection data used in this paper consists of approximately 84,000 defects collected from 1,084 turbines over 17 wind farms across North America (see Fig. 6).

Since the wind farm operators cannot afford to have every turbine inspected, data was collected for only a sample of the turbines in the fleet. Also, over time, different companies were hired for inspections making combining results difficult. By consulting with a turbine engineer, a common defect score was developed.

Identifying the specific defect signature of blade erosion is very important for accurate predictions. The inspection data contained records that did not match the erosion failure we were interested in. Cosmetic defects and damage caused by lightning strikes are considered noise as these do not cause performance degradation in the turbine. In addition to the location of defects along the length of the blade, the location along with width is also analyzed. Defects on the leading edge of the blade and located from the tip of the blade to half the length of the turbine blade. Figure 7 illustrates this by showing the distribution of defects that occur along the blade. This defect filtering is critical to accurately predicting leading-edge erosion.
4.2 Turbine Performance Data

Modern wind turbines have many moving parts in the nacelle and almost all of them have either a temperature or vibration sensor installed. Turbine controllers use this data to optimize performance, decrease output in high stress situations, or shut down for protection. The high resolution data is collected and stored by SCADA systems as 10 minute average or one minute snap shot records. The turbine data that is most useful in this analysis is power output and wind speed. These two measurements are key to determining if the turbine is not performing at expected efficiency. Other turbine data is used for fatigue analysis where time spent in certain high stress conditions can be aggregated. This allows for turbines to be compared in a snapshot in time with blade inspections.
4.3 Meteorological data

Weather data can be useful to calculate wind turbulence, wind sheer, precipitation, and temperature. Most wind farms have at least one met tower where these instruments are installed. For precipitation data, local airports or other National Weather Service data sources may be needed.

5 Solution Approach

At first glance the inspection data suggests using a time series model for this problem; however, time series data is a sequence of data points that measures the same metric over a consistent period of time. Data that is collected based on the occurrence of an event is not considered to be time series data. The turbine inspection data has sparse and inconsistent records on inspection and is based on when the turbine inspection was conducted.

To use a time series approach for our data, means employing one of the many interpolation formulas to convert the data into equally spaced observations. However, there are challenges to using this approach. The inspection data comes from several sites with different time resolution (some monthly and others quarterly). The usual approach to interpolate this type of data, is to sample at the smallest time resolution, which will end up with a time series that is sampled
each month, but is actually recorded every quarter. This can affect our accuracy negatively.

Instead of a time series approach, with data that has the granularity of the time stamp for when the inspection was done, a snapshot approach was employed. The data was grouped at the turbine level and for each turbine features engineered to contain aggregate of fatigue metrics over the turbine’s entire life.

When it comes to modeling data, we could not rely on the results of an individual model because of the commonality that individual models suffer far too much from bias and/or variances; thus, we decided to incorporate ensemble learning, which makes predictions based on several different models. Combining each model results in the ensemble model being more flexible which creates less bias and less variance. The model becomes less data-sensitive[11].

Two of the most popular methods ensemble learning are bagging and boosting. Bagging encompasses the training of several individual models in parallel where each model is trained by a random subset of data. Boosting requires the training of several individual models in sequence where each model learns from the errors made by the previous model[11]. In the case of prediction-based problems involving small-to-medium structured/tabular data, the preferred approach is utilizing decision-tree based algorithms[12]. XGBoost is a decision-tree-based algorithm which falls under the ensemble learning method that uses a gradient boosting framework[12].

6 Analysis

6.1 Feature Engineering

The most important feature to create is the blade tip speed. Since the data set consists of turbines of various blade lengths and rotational speeds, this feature will allow all turbines to have a comparable metric for modeling. Blade tip speed is calculated according to Equation 1

\[ v = \omega \times r, \]  

where \( v \) is the linear speed at the blade tip (m/s), \( \omega \) is the rotational speed (radians/minute), and \( r \) is the blade length (meters).

Figure 10 shows an example of tip speed plotted by time for one turbine. You can see the plateau of speed is reached when wind speed is above a threshold. This rated speed varies by turbine model. Many other features are also used including wind turbulence, wind sheer, and the product of blade tip speed and rainfall.

These features are aggregated over time between the date of inspection and the date of commissioning. These final aggregated features are the inputs to the final regression model to predict the turbine’s blade condition. An example of some of these features over time can be seen in figure 11.
6.2 Machine Learning Model

The goal of the machine learning model is to predict the severity score of turbines using the aggregated features created to capture blade fatigue. This model is a boosted random forest regressor with ten-fold cross validation. Figure 12 shows the individual turbine results colored by site.
The model residual analysis shows a well-defined normal distribution around zero as seen in Fig. 13. This suggests low bias overall but further analysis is needed to verify.
The residual bias in the RainWindLog feature was investigated in Fig. 14. There is no clear pattern to the residuals which is a good sign of low bias.

There does exist bias in the residual plot of severity shown in Fig. 15. There is a linear relationship for severity below 2.8 where very low severity shows an overestimated prediction and higher severity shows an underestimated prediction. At very high severity, there is a bias of underestimated predictions for most points above 3.3. The cause of this residual distribution is not clear.
The residual analysis by wind farm also shows non-uniform residuals where there is strong bias in both directions depending on the wind farm site. There are numerous potential reasons for this behaviour. First, each site has unique weather complexities that can skew the fatigue metrics. Second, the turbine performance data set used in this analysis is not complete for all wind farms. This could skew the aggregations and can be address with proper imputation. Third, the amount of blade damage varies by site and could cause, or be caused, by the residual by severity analysis discuss earlier. For example, K2 Wind has the largest severity values in the data set (Fig.12). This large underestimation of predicted scores matches both analysis.
6.3 Feature Importance

A benefit to modelling with gradient boosting is the fact that it is able to automatically provide estimates of features that contributed greatly to the model. A feature is considered important if modifying its values significantly changes the model’s predictability. As shown in Fig 12, Rainwindlog came out as a significant feature in blade erosion prediction. This feature accounts for the tip speed of the blade, the length of the turbine blade length which varies across location and the amount of rain. Rainwindlog coming out as the best feature was no surprise since blade erosion is mostly driven by the impact of water droplets on fast moving bladed. This suggests that being able calculate the amount of rain that a turbine received along with it blade length and rotation speed, blade erosion can predicted at 87 percent of the time. For the other 13 percent, the model of the turbine and the turbines age come into play when the Rainwindlog alone cannot be used to predict blade erosion.
7 Ethics

Our study attempts to make wind turbine farm operations more efficient and less expensive, which goes to explain our support for renewable energy. One of the many benefits of wind energy is that it produces no greenhouse emissions. However, there are many debates surrounding its benefits. There are concerns about the environmental and economic effects of wind turbines such as its effect on landscape, animal habitat and the economy of the towns where they are built[13].

For wind turbines to be effective they have to be built in high wind areas. These include mountain tops and ridges. However, people believe turbines built in these areas ruin the landscape and causes economic loss. A town in Scotland called Loch Ness where millions of people travel each year to see its hills and mountains is an example [14]. In 2015 plans were made to develop 30 miles of wind farms. People protested against the idea claiming that the environment will lose its natural beauty which will negatively affect the economy of the town. In addition to that, people believe that this will introduce industrialization in a place that has been specifically reserved for nature. On the topic of economical effect some have also argued that besides the large amount of money that needs to be invested in setting up a wind farm, studies show that Wind Turbines Farms built near habitats have caused homes to lose their economic value [13].

Presently there is no way of guaranteeing the amount of wind in a given day. To make up for this unpredictable nature, more wind turbines are built to make up for low wind days. More wind turbines also mean that much more materials are needed to produce the turbine. That is, more material per energy produced. These materials include chemical compound used for making turbine blade, rare earth metals to create magnets as well as all the steel and concrete needed to build the turbines. Most of these materials have to be mined from the earth. In
addition, because of the intermittent nature of wind, power systems that run on wind energy usually have a backup power supply which is usually fossil fuel. For this reason Germany, a world leader in renewable energy, has not been able to cut down its carbon emission significantly even after high investments in wind and solar energy [15].

Wind Turbine farms have also been known to negatively impact wildlife. There is evidence of bats and birds colliding with turbine blades according to a study by the National Wind Coordinating Committee (NWCC)[16]. Even though the impact is low, the study has evidence of birds and bat deaths from collisions with wind turbines due to changes in air pressure caused by the spinning turbines. The aerodynamic nature of the blades causes the eardrums of bats to explode/implode which causes them to lose orientation and fall. The good news here is that there is ongoing research to reduce bird and bat deaths caused by Wind Turbines. Researchers have found that bats are most active when wind speeds are low, so keeping wind turbines off during times of low wind speeds can possibly reduce bat deaths by more than half without significantly affecting power production [16].

8 Conclusion

The effect of rain at high blade tip speed as a significant driving force for blade erosion. This paper confirms this theory and is in agreement with the scientific consensus on the issue[17]. This work offers a methodology for using turbine historical fatigue metrics and machine learning to predict blade remaining lifetime. These techniques can be applied to many predictive maintenance applications in wind turbines and beyond.

9 Future Work

According to the literature[17] there are many particulates that have been found to contribute to erosion. Data collected on dust, sand, crop debris, and insects could server to improve model accuracy. There is available data on severe thunderstorms[18] that can help improve the fatigue features and quantify hail impacts to the blade. Lightning damage is a source of noise for predicting leading edge erosion. Since leading-edge erosion can impact the aerodynamics of the blade, there is an overall performance impact to the turbine. This can be quantified with machine learning using the first year of life as a baseline. This under-performance estimation can be used as a feature input to improve model accuracy.

References


8. L.Yue H.Kim, C. Kim. A study of the growth of single-phase mg0. 8.


Appendix

Code Repository

The code used for this paper can be found at the GitHub repository below. https://github.com/CaseyRMartinez/SMU_Capstone
Other Resources

https://www.kaggle.com/novinsh/surprise-me-2-with-lstm
https://stats.stackexchange.com/questions/354019/lstm-time-series-classification-using-keras
https://www.energy.gov/eere/wind/articles/top-10-things-you-didnt-know-about-wind-power
http://www.windfarmbop.com/category/met-mast/