Engineering Applications of Artificial Intelligence 55 (2016) 331–338

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Data-driven multivariate power curve modeling of offshore wind turbines

Olivier Janssens [a,*], Nymfa Noppe [b], Christof Devriendt [b], Rik Van de Walle [a], Sofie Van Hoecke [a]

[a] Data Science Lab, Department of Electronics and Information Systems, Ghent University - iMinds Sint-Pietersnieuwstraat 41, 9050 Ghent, Belgium
[b] Offshore Wind Infrastructure Lab and the Brussels Wind Energy Research Institute, Department of Mechanical Engineering, Vrije Universiteit Brussel, B-1050 Brussels, Belgium

ARTICLE INFO

Article history:
Received 21 September 2015
Received in revised form
26 July 2016
Accepted 5 August 2016
Available online 18 August 2016

Keywords:
Performance monitoring
Condition monitoring
Machine learning
Data mining
Wind energy

ABSTRACT

Performance monitoring of offshore wind turbines is an essential first step in the condition monitoring process. This paper provides three novelties regarding power curve modeling. The first consists of illustrating that univariate power curve modeling can be improved by the use of non-parametric methods such as stochastic gradient boosted regression trees, extremely randomized forest, random forest, K-nearest neighbors, and the method of bins according to the IEC standard 61,400–12-1. This is confirmed on both a synthetic data set and a real live data set containing data from three offshore wind turbines. The second novelty consists of an improvement regarding overall power curve modeling results by the use of multivariate models which incorporate the wind direction, rotations per minute of the rotor, yaw, wind direction and pitch additional to the wind speed. The best improvement is achieved by the stochastic gradient boosted regression trees method for which the mean absolute error can be decreased by up to 27.66%. The third novelty consists of making a synthetic data set available for bench-marking purposes.

1. Introduction

The significance of renewable energy has increased and will keep increasing over the coming decades as the European Union has set a 20% renewable energy target for 2020. With this goal in mind, the European Wind Energy Association (EWEA) proposed three possible growth scenarios for wind energy towards 2020. The three scenarios project that there will be an increase in installed wind turbine capacity of 41%, 64%, or 84,9% respectively compared to 2013 (EWEA, 2014).

To make wind energy competitive with non-renewable energy sources, the unavailability of wind turbines due to maintenance needs to be reduced. This is especially challenging for offshore wind turbines as maintenance is not always possible because of the weather (Daneshi-Far et al., 2010). A cost efficient maintenance strategy is predictive maintenance (PM). PM employs condition monitoring (CM) and structural health monitoring (SHM) to prematurely detect component faults. Early fault detection enables maintenance operations to be scheduled efficiently, therefore reducing the downtime and energy loss of the wind turbines and the resulting costs.

A comprehensive review by Hameed et al. (2009) illustrates that several CM and SHM topics have emerged, such as oil analysis for pressure loss detection or contamination detection, blade strain monitoring for crack detection and acoustic emission/vibration analysis for bearing and gear fault detection in the drivetrain Janssens et al. (2016). Most of these techniques are designed to detect specific faults or anomalies and require additional sensors to be mounted in or on the turbine. Nevertheless, the failure of a critical component results in the performance of the turbine to deviate significantly from its expected performance. Hence, performance monitoring is an important first step in condition monitoring. Performance monitoring is done using power curves which express the relationship between the wind speed and output power of a turbine as can be seen in Fig. 1. The goal of a power curve model is to capture the wind turbine's behavior in normal operational conditions. Hence, by providing the wind speed to the power curve model, the expected output power is calculated. If the actual measured output power deviates too much from this power production prediction, further investigations can be required.

A theoretical power curve, between the cut-in speed and the rated output speed, of a wind turbine is given by Eq. (1), where P is

* Corresponding author.
E-mail addresses: odjansse.janssens@ugent.be (O. Janssens), nymfa.noppe@vub.ac.be (N. Noppe), christof.devriendt@vub.ac.be (C. Devriendt), rik.vandewalle@ugent.be (R.V.d. Walle), sofie.vanhoecke@ugent.be (S.V. Hoecke).
URL: http://www.datasciencelab.ugent.be/ (O. Janssens).

http://dx.doi.org/10.1016/j.engappai.2016.08.003
0952-1976/

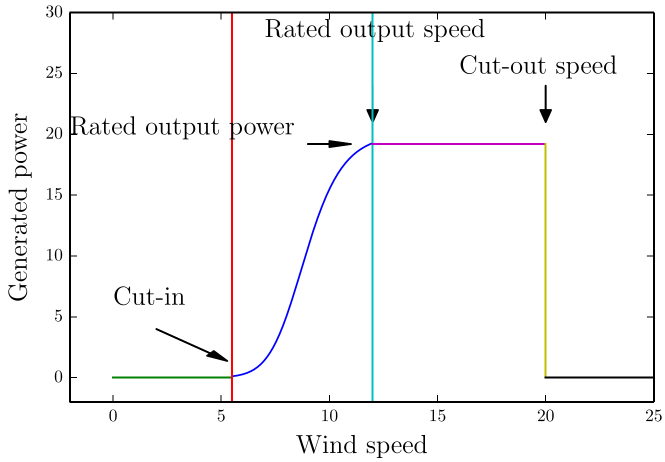

Fig. 1. Example of a wind turbine power curve. The minimum speed at which a wind turbine produces power is known as the cut-in speed. The rated output power is the maximum output power and the cut-out speed is the maximum operational speed.

the output power, ρ the air density, u the wind speed, A the swept rotor area and C_p the power coefficient. In this equation the independent variable is u, while the other variables are taken fixed. However, a power curve is influenced by more than the wind speed. For example, the air pressure varies by about 10% depending on the weather phenomena (Schlechtingen et al., 2013b). Other phenomena also influence the power curve such as the wind field, yaw and pitch misalignment, shading effects by nearby obstacles, turbulent air created by turbines nearby and the mechanical behavior of the turbines.

$$P = 0.5\rho A C_p u^3 \quad (1)$$

A theoretical model does not consider these complicated interactions. Nevertheless, it is possible to take these into account by using data-driven approaches wherein a power curve is constructed directly from the data instead of using the underlying physics. By using a data-driven methodology no explanation or origin is linked directly to certain phenomena. A model is created based on measurements, hence the phenomena present in the data are encompassed in the model. For example, if due to the wind field the turbine underperforms at high wind speeds, a data-driven model will take this into account because this phenomenon is manifested in the data. By being able to take the phenomena which are present in the data into account, a more accurate power prediction is possible. Furthermore, by being agnostic to external influences, which contribute to the generated power besides the wind speed, only data from the turbine's supervisory control and data acquisition (SCADA) system is required making power curve based performance monitoring a suitable and cheap first step for condition monitoring.

The remainder of this paper is as follows. First, Section 2 discusses related work. Section 3 elaborates on the different techniques employed in this paper. While, Section 4 describes the data sets used for the performance study. Section 5 discusses the univariate power curve modeling results and their analysis. The comparison indicates a performance limit which is expected as this technique only uses the information contained in the wind speed. In order to overcome this problem more information, which is also readily available in the SCADA system, is subsequently incorporated in the models by the use of multiple variables. Therefore, Section 6 discusses the results and accompanying analysis of multivariate power curve modeling. Finally, in Section 7, a conclusion is provided together with a short summary of future steps regarding performance monitoring.

2. Related literature

To use a power curve for condition monitoring purposes, first the power curve has to be modeled using data gathered during normal operational conditions. Afterwards, the actual amount of generated power can be compared to the predicted output of the power curve to detect anomalies. As the model of a power curve in a healthy condition is crucial to this approach, many modeling techniques have been proposed and compared in the recent years.

2.1. Single-sensor based models

A power curve uses one parameter as input, hence, only data from one sensor is required together with the corresponding generated power to construct a power curve. In recent years the focus has been on modeling the relation between these two parameters and to improve the modeling approaches. For example, Kusiak et al. (2009) compare two parametric models with five non-parametric models on separate data sets. The two parametric models used are both four-parameter logistic curves, one for which the parameters are optimized by maximum likelihood and one for which they are optimized by the least-squares method. The non-parametric models are a multi-layer perceptron, a random forest, a M5P decision tree, a boosting tree and a K-nearest neighbors (KNN) regression model. From the non-parametric models, the KNN model achieves the best results. From the parametric models, the four parametric logistic curve optimized with least squares performs the best.

A recently published comprehensive review on power curve modeling by Lydia et al. (2014) also presents a comparative literature study on parametric and non-parametric methods. Methods reviewed are, among others, linearized segmented model, polynomial curve fitting, both four and five parameter logistic curves fitted using several different optimization algorithms, neural networks (NNs) and fuzzy methods. When comparing the two reviews it can be concluded that by fitting a five-parameter logistic (5PL) curve using differential evolution (DE) the best results are achieved.

2.2. Multi-sensor based models

It was recently shown that other possible combinations of parameters, which are measured and stored by the SCADA system, can be used for performance monitoring of a wind turbine. In recent work of Kusiak and Verma (2013) the rotor curve (mapping between rotor speed and wind speed) and blade pitch curve (mapping between the turbine's pitch and wind speed) are used for more precise anomaly detection.

In the work of Schlechtingen et al. (2013a,b) methods such as cluster center fuzzy logic, NN, KNN and adaptive neuro-fuzzy interference system models are compared for power curve monitoring in order to construct power curves. The results indicate that when augmenting the wind-speed based models by using the ambient temperature and wind direction as additional inputs, the variance in the generated power is better accounted for. Because of this, better prediction results are achieved and more accurate anomaly detection is possible.

2.3. Datasets

Due to the confidential nature of wind turbine data, no publicly made available data set exists for bench-marking purposes. Nevertheless, recently a synthetic data set generation procedure was proposed by Jin and Tian (2010) and further detailed by Lydia et al. (2013).