Energy Conversion and Management 171 (2018) 156–169

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Implementation of novel hybrid approaches for power curve modeling of wind turbines

Mehmet Yesilbudak

Nevsehir Haci Bektas Veli University, Faculty of Engineering and Architecture, Department of Electrical and Electronics Engineering, 50300 Nevsehir, Turkey

ARTICLE INFO

Keywords:
Wind turbine
Power curve
Clustering
Filtering
Modeling
Accuracy

ABSTRACT

In wind energy conversion systems, a power curve links the wind speed to the power produced by a wind turbine and an accurate power curve model helps wind power providers to capture the performance of wind turbines. For this purpose, this paper presents the implementation of novel hybrid approaches to the power curve modeling process of wind turbines. As a result of employing the complementary phases called clustering, filtering and modeling in this process, the k-means-based Smoothing Spline hybrid model achieves the most accurate power curve in terms of sum of squared errors, coefficient of determination and root mean squared error. On the other hand, the k-medoids + +-based Gaussian hybrid model causes the most inconsistent power curve in terms of the mentioned goodness-of-fit statistics. Furthermore, all of hybrid power curve models constructed in this paper outperform the conventional linear, quadratic, cubic, exponential and logarithmic benchmark models with the high improvement percentages. Finally, the proposed hybrid power curve models are shown not to be dependent on the initial raw power curve data.

1. Introduction

With the population growth and economic development, the demand for electricity is steadily increasing in the world. Due to the rapid depletion and harmful emissions of fossil fuels, there is a growing trend for installing solar photovoltaic, wind and biomass power plants in order to prevent the possible energy crisis in the future [1,2]. Especially, wind power was found to be the world's second largest annual market with the capacity addition of 55 GW in 2016 [3]. In wind energy conversion systems, wind turbine power curves play a significant role in wind turbine selection, wind energy assessment, condition monitoring and troubleshooting, and predictive control and optimization [4]. However, the power curve modeling of wind turbines is still a compelling task owing to the operating conditions on site different from the conditions under which wind turbines were calibrated [5]. In addition, the empirical power curves are affected from environmental and topographical conditions, maintenance and repairment, control system issues, incorrect controller settings, sensor malfunctions, blade pitch angle errors and blade damage [6]. For these reasons, it is needed to model the wind turbine power curves in the form of reflecting the normal turbine behavior.

Many different parametric and non-parametric methods have been used in the literature. 6, 5, 4 and 3-parameter logistic functions were applied for the power curve modeling of different wind turbines [7]. 5 and 3-parameter logistic functions were recommended according to the intended use. Differential evolution-based 5-parameter logistic function led to the encouraging modeling in terms of mean absolute error [8]. The combination of nonlinear autoregressive model with exogenous variables and differential evolution-based 5-parameter logistic function performed well in terms of root mean squared error and coefficient of determination [9]. Modified hyperbolic tangent function, sixth-order polynomial function and three-parameter exponential function were compared for the power curve approximation of a wind turbine [10]. Three-parameter exponential function provided the lowest value of mean absolute percentage error. Backtracking search, cuckoo search and particle swarm optimization-based modified hyperbolic tangent functions showed the promising approximation in terms of root mean squared error [11]. Penalized spline regression outperformed polynomial regression, locally weighted polynomial regression and spline regression methods in terms of normalized mean absolute percentage error and root mean square error [12]. 5th and 9th-order polynomial functions, double exponential and logistic functions, k-nearest neighbor regression and multilayer perceptron were employed for the site-specific characterization of wind turbine power curves [13]. Multilayer perceptron demonstrated the lower levels of mean error and mean absolute error.

Cubic spline interpolation and least square methodology gave accurate mathematical modeling of the wind turbines having smooth power curve [14]. Spline kernel function-based support vector machine algorithm reflected the dynamic properties of a power curve having

E-mail address: myesilbudak@nevsehir.edu.tr.

https://doi.org/10.1016/j.enconman.2018.05.092
Received 14 March 2018; Received in revised form 3 May 2018; Accepted 25 May 2018

equal-size wind speed partitions at a low computational cost [15]. Differential evolution-based 5-parameter logistic function, k-nearest neighbor regression, random forest regression, extremely randomized trees and stochastic gradient boosted regression trees were utilized for the data-driven power curve fitting of different wind turbines [16]. The best fitting was achieved by the stochastic gradient boosted regression trees in terms of mean absolute error. Generalized mapping regressor, feed-forward multilayer perceptron and general regression neural network methods were proposed for modeling the relationship between wind power and wind speed [5]. The best models were found as general regression neural network and generalized mapping regressor without interpolation in terms of absolute error and symmetrical absolute percentage error, respectively. Cluster center fuzzy logic modeling produced the smaller root mean squared error than least square polynomial methodology [17]. Adaptive neuro-fuzzy inference system, neural network, cluster center fuzzy logic and k-nearest neighbor models were built for monitoring the wind turbine power curve [18]. The adaptive neuro-fuzzy inference system accomplished the qualified monitoring performance in terms of mean absolute error, mean absolute percentage error and root mean squared error. In addition to these studies, there are many other methods used for the power curve modeling of wind turbines in the literature, such as approximate cubic function [19], copula method [20], linearized segmented model [4], stochastic modeling [21], least median of squares methodology [22] and extreme function theory [23].

The majority of the literature has focused on different parametric and non-parametric modeling of wind turbine power curves. However, most of them ignore to divide the power curve modeling process into simple, useful and efficient phases in order to capture the wind turbine's behavior in normal operational conditions. For this purpose, the main motivation of this study is to constitute the power curve modeling process with the clustering, filtering and modeling phases. In this framework, the main novelties of this study lie in the employment of k-means, k-means++, k-medoids and k-medoids++ clustering techniques with Squared Euclidean, City Block and Cosine distance measures in the power curve clustering phase, in the application of multivariate outlier detection approach based on Mahalanobis distance and chi-square cumulative distribution to each created cluster in the power curve filtering phase, and finally in the implementation of Polynomial, Fourier, Gaussian, Sum of Sines and Smoothing Spline curve fitting methods on the refined power curves in the power curve modeling phase. In addition, all of hybrid power curve constructions are compared in terms of sum of squared errors (*SSE*), coefficient of determination (R^2) and root mean squared error (*RMSE*) in detail. As a result, the proposed hybrid methodologies demonstrate the promising performance for the wind turbine power curve modeling.

The remainder of this paper is organized as follows: Section 2 describes the mathematical background of the proposed hybrid power curve modeling approaches. Section 3 elaborates the clustering, filtering and modeling phases applied to wind turbine power curves. Finally, Section 4 states the conclusions together with future works.

2. The proposed hybrid power curve modeling approaches

In this subsection, the algorithmic procedure of the proposed hybrid power curve modeling approaches is explained in detail. The hybrid power curve modeling process is constituted with three main phases called power curve clustering, power curve filtering and power curve modeling in order to create very well-suited power curve models.

In the power curve clustering phase, the well-known partitioning methods k-means and k-medoids and their commonly-used variants k-means++ and k-medoids++ are employed for the highly-correlated segmentation of power curve data. k-means and k-medoids algorithms partition a set of n objects into k clusters with the purposes of increasing intracluster similarity and decreasing intercluster similarity. In the k-means algorithm, the similarity of each cluster is measured regarding the mean value of all objects in that cluster as a reference point. In the k-medoids algorithm, the dissimilarity of each cluster is determined considering the medoid of all objects in that cluster as a representative object. The partitioning process iterates until the convergence of square-error function in the k-means algorithm and until the minimization of absolute-error function in the k-medoids algorithm. The square-error criterion (E_1) and the absolute-error criterion (E_2) are defined as below [24,25], where k is the number of clusters, p is the point representing an object in cluster C_i or C_j, m_i is the mean of cluster C_i and o_j is the representative object of cluster C_j. The value of k is assigned as 10 in this study.

Table 1
Descriptive statistics of W_P and W_S parameters.

Statistics	Parameters	
	W_P (kW)	W_S (m/s)
Maximum value	2030.20	19.92
Mean value	688.286	9.252
Minimum value	0.23	5
Standard deviation	608.4436	3.1269

$$E_1 = \sum_{i=1}^{k} \sum_{p \in C_i} |p - m_i|^2 \quad (1)$$

$$E_2 = \sum_{j=1}^{k} \sum_{p \in C_j} |p - o_j| \quad (2)$$

k-means++ and k-medoids++ algorithms are identical to k-means and k-medoids algorithms, respectively, but differs in the initial centroids/medoids setting. Instead of choosing the initial centroids/medoids arbitrarily, k-means++ and k-medoids++ algorithms compute the probability of how well a given point is doing acting as a possible centroid/medoid [26]. The weighted probability distributions, WPD_1 for the cluster centroid initialization in k-means++ algorithm and WPD_2 for the cluster medoid initialization in k-medoids++ algorithm are defined as below [27], where $d(p, m_i)$ denotes the distance between p and m_i and $d(p, o_j)$ denotes the distance between p and o_j.

$$WPD_1 = \frac{d^2(p, m_i)}{\sum_{h \in C_i} d^2(h, m_i)} \quad (3)$$

$$WPD_2 = \frac{d^2(p, o_j)}{\sum_{h \in C_j} d^2(h, o_j)} \quad (4)$$

The mentioned partitioning methods compute centroids/medoids differently for the different distance measures. For this reason, Squared Euclidean, City Block and Cosine distance measures are employed for uncovering their effects on the nearest neighborhood recovery. These distance measures are expressed as below [28,29], where X is the data matrix with m objects and n dimensions, x_i is the ith row of X and x_j is the jth row of X. In addition, the silhouette coefficient is utilized for measuring the quality of clustering solutions. The silhouette value $s(i)$ of ith data point is expressed as below [30], where $a(i)$ represents the average distance from ith data point to other data points in the same cluster and $b(i)$ represents the minimum average distance from ith data point to the points in other clusters. It should be noted that a silhouette value close to 1 indicates that the corresponding data point is well-matched to its own cluster and poorly-matched to other clusters.

$$d_{SEuclidean}(x_i, x_j) = (x_i - x_j)(x_i - x_j)' \quad (5)$$

$$d_{CBlock}(x_i, x_j) = \sum_{t=1}^{n} |x_{it} - x_{jt}| \quad (6)$$