Wind Turbine Power Curve Modeling Using Advanced Parametric and Nonparametric Methods

Shahab Shokrzadeh, *Student Member, IEEE,* Mohammad Jafari Jozani, and Eric Bibeau

***Abstract*—Wind turbine power curve modeling is an important tool in turbine performance monitoring and power forecasting. There are several statistical techniques to fit the empirical power curve of a wind turbine, which can be classified into parametric and nonparametric methods. In this paper, we study four of these methods to estimate the wind turbine power curve. Polynomial regression is studied as the benchmark parametric model, and issues associated with this technique are discussed. We then introduce the locally weighted polynomial regression method, and show its advantages over the polynomial regression. Also, the spline regression method is examined to achieve more flexibility for fitting the power curve. Finally, we develop a penalized spline regression model to address the issues of choosing the number and location of knots in the spline regression. The performance of the presented methods is evaluated using two simulated data sets as well as an actual operational power data of a wind farm in North America.**

***Index Terms*—Nonparametric regression, penalized spline regression, polynomial regression, wind energy, wind turbine power curve.**

I. Introduction

RENEWABLE energy plays an important role in addressing global energy and environmental challenges. To improve energy sustainability and to mitigate risks from a business-as-usual approach, large-scale deployment of renewable sources has significantly increased over the last decade [1]. Among renewable energy technologies, wind energy has been the fastest growing source in electricity generation [2]. Higher share of wind produced electricity in the energy sector motivates the analysis of the performance of wind power generators. Uncertainties and deviations in the generated output power can cause serious challenges in the energy management systems (EMS) and impact the reliability of the power grid [3]. Therefore, effective integration of wind power into the power systems requires accurate estimation of the turbine power curve for operational management of wind energy as well as performance monitoring of turbines [4]–[6]. Moreover, accurate estimation of wind turbine power curve is required to more realistically size the storage capacity for wind energy integration [7].

The power curve of a wind turbine presents the electrical power output ratings of the machine for different wind speeds [8]. A typical wind turbine power curve has three main characteristic speeds: 1) cut-in (V_c); 2) rated (V_r); and 3) cut-out (V_s) speeds. The turbine starts generating power when the wind speed reaches the cut-in value. The rated speed is the wind speed at which the generator is producing the machine's rated power. When the wind speed reaches the cut-out speed, the power generation is shut down to prevent defects and damages [9]. Theoretical power curves are supplied by manufacturers assuming ideal meteorological and topographical conditions. In practice, however, wind turbines are never used under ideal conditions, and the empirical power curves could be substantially different from the theoretical ones due to the location of the turbine, air density, wind velocity distribution, wind direction, mechanical and control issues, as well as uncertainties in measurements.

There are several statistical methods to fit the empirical power curve of a wind turbine [10]. These methods can be classified into parametric and nonparametric techniques [11], [12]. Parametric techniques are based on mathematical models that are often built by a family of functions with a number of parameters to describe the turbine power curve [13]. Examples of these models include segmented linear models [14], polynomial regression [15], [16], and models based on probabilistic distributions such as four- or five-parameter logistic distributions [12], [17]. Parametric methods are usually restricted by their nature. Unlike parametric techniques, nonparametric methods do not impose any prespecified model and attempt to produce an estimate of the power curve that is as close as possible to the observed data subject to the smoothness of the fit. Such methods have major advantages over parametric methods as they can accurately model a wide range of possible shapes of power curves. Examples of nonparametric techniques include neural networks (e.g., generalized mapping regressor and feed–forward multi layer perceptron [18]), fuzzy logic methods (e.g., fuzzy cluster centre models [19]), and data mining methods (e.g., the multilayer perception, the random forest, and the *k*-nearest neighbor [20]).

No one model fitting approach dominates all others over all possible observations obtained from different wind turbines. On a particular data set, a specific method might work best, but on other data sets, other methods might be more applicable. Therefore, it is important to investigate the performance of

Manuscript received March 13, 2014; revised June 12, 2014; accepted July 12, 2014. Date of publication September 08, 2014; date of current version September 16, 2014. This work was supported by the NSERC/Manitoba Hydro Industrial Research Chair in Alternative Energy.

S. Shokrzadeh and E. Bibeau are with the Department of Mechanical Engineering, University of Manitoba, Winnipeg, MB R3T 2N2 Canada (e-mail: Shahab_Shokrzadeh@umanitoba.ca; Eric.Bibeau@umanitoba.ca).

M. Jafari Jozani is with the Department of Statistics, University of Manitoba, Winnipeg, MB R3T 2N2 Canada (e-mail: M_Jafari_Jozani@umanitoba.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSTE.2014.2345059

different statistical procedures for power curve fitting and decide which method produces better results for a given data set. In this paper, we focus on wind turbine power curve modeling based on available operational output power data using four parametric and nonparametric methods. We present polynomial regression as a benchmark parametric method for power curve fitting. This method used extensively in the literature, however, suffers from its global nature and sensitivity to anomalies within observations. In this method, one usually needs a high-degree polynomial regression model to provide a good fit to the observed data set. Fitting a high-degree polynomial regression model results in a good fit to the observed data set but may overfit the data points [21], and the fitted power curve will closely follow the noise of the power generating system. To avoid such problems, we propose using the locally weighted polynomial regression as a nonparametric method and study some of its properties on simulated as well as real data sets. Cubic spline regression is another nonparametric method that has been introduced for wind turbine power curve fitting [22], [23]. However, there are a number of practical issues with this method such as choosing the number and the location of knots to fit the cubic spline model. In addition, while these models perform well for wind turbines with smooth power curves, their performance outside the boundary knots could be undesirable. Natural cubic spline regression is proposed to improve the performance of the cubic spline regression models outside the boundary knots. Finally, we develop a penalized spline regression model that provides an enhanced performance compared with the spline regression by addressing the challenge of choosing the number and the location of knots. We note that the nonparametric methods developed in this paper are more flexible, less sensitive to anomalies within observations, easier to implement, and computationally more feasible compared with other methods in the literature. While our proposed methods show promising results for modeling the wind power generation, one might also be able to use our results to obtain efficient and easy-to-implement methods for characterizing wind turbine power curves, which can be used in other applications such as wind power forecasting and on-line monitoring of power curves for detecting anomalies in a wind turbine power generation process.

The outline of this paper is as follows. In Section II, we first discuss different parametric and nonparametric power curve estimation methods by introducing two typical power curves and simulating two random data sets with normal errors. We then present the theoretical foundation of each method and apply them on the simulated data sets. In Section III, we present actual operational data sets of a wind farm in North America and investigate the performance of each method. The results of the proposed techniques and the evaluation metrics are also presented. Section IV provides concluding remarks and future works.

II. Power Curve Estimation

The theoretical wind power available from the mass flow rate of air through the turbine blades swept area is obtained by

$$P_w = \frac{1}{2}\,\rho\, A\, v^3 \tag{1}$$

where P_w is the wind power in W, ρ is the air density in kg/m^3, A is the turbine rotor area in m^2, and v is the wind speed in m/s [9]. The electrical power extracted by wind generators is

$$P_e = \eta\, C_p\, P_w \tag{2}$$

where C_p is the dimensionless power coefficient representing the theoretical amount of mechanical power that can be extracted by the turbine rotor, and η is the machine's overall efficiency [24], [25]. The power coefficient is a function of turbine blade tip speed ratio λ and the blade pitch angle θ [26]. The maximum theoretical mechanical power that can be extracted by wind turbines is 0.5926 and is known as the Betz limit [27].

In this section, we characterize the machine's power curve based on actual generated power data using four parametric and nonparametric methods. Wind speed and power data sets (v_i, p_i) are simulated from the additive model $p_i = f(v_i) + \epsilon_i$ with contaminated noise, where $f(\cdot)$ represents the manufacturer power curve. We consider two wind turbines to represent two different typical shapes of the power curve: wind turbine model V82 and model FL-255 with the rated capacity of 1650 kW and 250 kW, respectively, manufactured by Vestas Wind Systems A/S and Furlander AG. Fig. 1 shows theoretical power curves for these turbines with the scatter plots of simulated observations from each power curve. For each case, the observed wind power at a given wind speed is generated from a normal distribution with the mean equal to the manufacturer power curve and a constant standard deviation σ_ϵ, where $\sigma_\epsilon = 100$ for turbine model V82 and $\sigma_\epsilon = 20$ for turbine model FL-255. Each data set consists of a total number of 720 pairs of observations, denoted by (v_i, p_i), $i = 1, \ldots, n = 720$. The wind speed data are generated from a Weibull distribution representing the hourly wind distribution of the wind farm studied in Section III.

A. Polynomial Regression

Polynomial regression has been extensively used in the literature to estimate the power curve of wind turbines. This model can be considered as a standard extension of the linear regression $p_i = \beta_0 + \beta_1 v_i + \epsilon_i$, with a polynomial function

$$p_i = \beta_0 + \beta_1 v_i + \beta_2 v_i^2 + \cdots + \beta_k v_i^k + \epsilon_i. \tag{3}$$

Model (3) can be written as

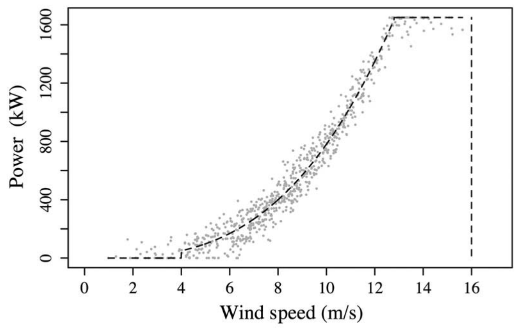
$$\mathbf{P} = \mathbf{V}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{4}$$

where $\mathbf{P} = (p_1, p_2, \ldots, p_n)^\top$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)^\top$, $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^\top$, and $\mathbf{V}$ is a matrix with its ith row being defined as $\mathbf{V}_i = (1, v_i, v_i^2, \ldots, v_i^k)$. We use the least squares method to estimate $\boldsymbol{\beta}$ by minimizing the residual sum of squares (RSS)

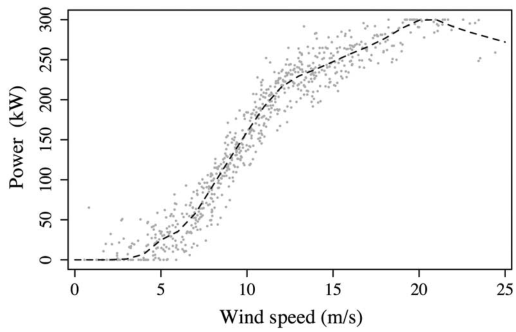
$$\text{RSS}(\boldsymbol{\beta}) = (\mathbf{P} - \mathbf{V}\boldsymbol{\beta})^\top(\mathbf{P} - \mathbf{V}\boldsymbol{\beta}). \tag{5}$$

Differentiating (5) with respect to $\boldsymbol{\beta}$, we solve

$$\frac{\partial \text{RSS}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\mathbf{V}^\top(\mathbf{P} - \mathbf{V}\boldsymbol{\beta}) = 0 \tag{6}$$

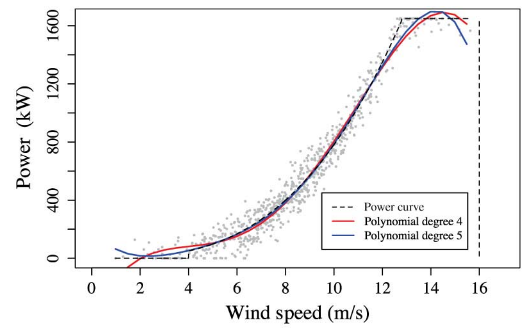


(a)

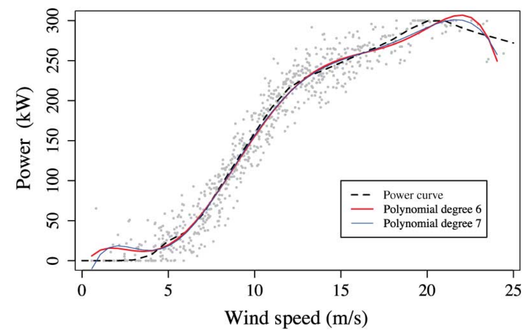


(b)

Fig. 1. The manufacturer wind turbine power curves for (a) turbine model V82 and (b) turbine model FL-255 with the scatter plots of 720 generated wind speed and hourly produced power for 1 month, assuming normally distributed noises about the manufacturer power curves with standard deviations $\sigma_\epsilon = 100$ and $\sigma_\epsilon = 20$, respectively.



(a)



(b)

Fig. 2. Fitted polynomial regression models for (a) turbine model V82 with degrees 4 and 5 and (b) turbine model FL-255 with degrees 6 and 7. The black-dashed curves are theoretical power curves from the manufacturers.

where $\mathbf{V}^\top$ stands for the transpose of $\mathbf{V}$ and obtain

$$\widehat{\boldsymbol{\beta}} = (\mathbf{V}^\top \mathbf{V})^{-1} \mathbf{V}^\top \mathbf{P}. \tag{7}$$

The fitted power curve at a specific wind speed value v_i is $\widehat{f}(v_i) = \mathbf{V}_i \widehat{\boldsymbol{\beta}}$. To obtain the degree of the polynomial regression, we use a 10-fold cross-validation by randomly dividing the observations into ten-folds of approximately equal sizes. Each time, a group of observation is considered as a validation set, and the remaining groups are used for the purpose of training a polynomial model of a specific degree for estimating the power curve. The tenfold cross-validation is then computed by

$$\text{CV}_{(10)} = \frac{1}{10} \sum_{i=1}^{10} \text{MSPE}_i \tag{8}$$

where MSPE_i is the mean-squared prediction error rate associated with the ith test group. Fig. 2 shows the fitted polynomial regression models for turbine model V82 with degrees 4 and 5, and turbine FL-255 with degrees 6 and 7. We use one standard error rule in conjunctions with cross-validation and choose a model with an error no more than one standard error above the error of the best model [21]. Here, to prevent overfitting the scatter plots, polynomial regression models with $k = 5$ and $k = 6$ are used for modeling the power curves of turbines V82 and FL-255, respectively.

B. Locally Weighted Polynomial Regression

Polynomial regression is limited by its global nature, where the fitted value of power at a given speed v_0 depends strongly on all data values even those v_is that are far from v_0. Also, it is not easy to achieve a functional form in a specific wind speed region without sacrificing the goodness of the fitted curve in other regions. In addition, polynomials are more sensitive to anomalies within the data. One way to avoid such problems is to fit a local regression model at a target point v_0. Locally weighted kth order polynomial regression model solves a separate weighted least squares problem at each target wind v_0 by finding $\widehat{\boldsymbol{\beta}}(v_0)$ as follows:

$$\widehat{\boldsymbol{\beta}}(v_0) = \arg\min_{\boldsymbol{\beta}} (\mathbf{P} - \mathbf{V}\boldsymbol{\beta})^\top \mathbf{W}_s(v_0)(\mathbf{P} - \mathbf{V}\boldsymbol{\beta}) \tag{9}$$

where $\mathbf{W}_s(v_0) = \text{diag}(\mathcal{K}_s(v_0, v_1), \ldots, \mathcal{K}_s(v_0, v_n))$ is a diagonal matrix, and $\mathcal{K}_s(v_0, v_i)$ is the smoothing kernel function. Using $\mathcal{K}_s(v_0, v_i)$, data points nearest to v_0 are given the highest weight and those farther away are given lower weights. This method is resistant against outliers by assigning low weights to observations, which generate large residuals [28]. For this analysis, we use the tri-cube kernel function

$$\mathcal{K}_s(v_0, v_i) = \begin{cases} \left(1 - \left|\frac{v_i - v_0}{s}\right|^3\right)^3, & \text{if } |v_i - v_0| \le s \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

However, one can also use other kernel functions such as the Gaussian kernel function. We can compute $\widehat{\boldsymbol{\beta}}(v_0)$ for each v_0 as follows:

$$\widehat{\boldsymbol{\beta}}(v_0) = (\mathbf{V}^\top \mathbf{W}_s(v_0) \mathbf{V})^{-1} \mathbf{V}^\top \mathbf{W}_s(v_0) \mathbf{P}. \tag{11}$$

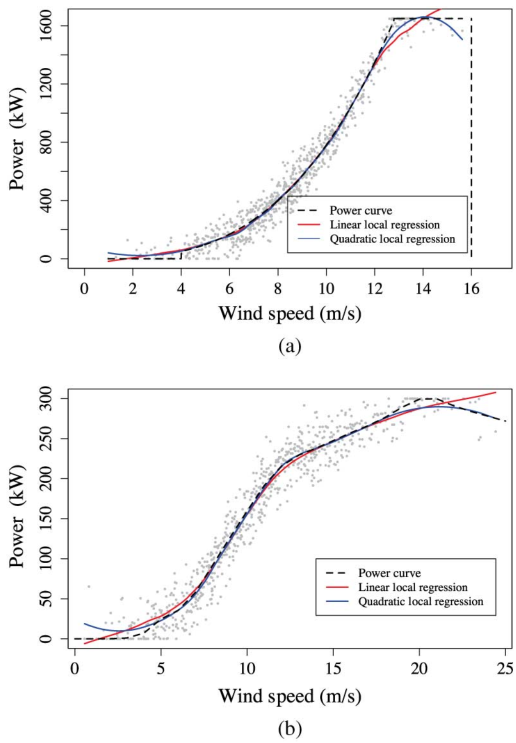


Fig. 3. Locally weighted linear and quadratic regression models for the power data sets generated from the theoretical power curves of turbines V82 (a) and FL-255 (b). The black-dashed curves are theoretical power curves from the manufacturers.

Also, the corresponding estimated power at wind speed v_0 is

$$\widehat{f}(v_0) = \mathbf{V}_0(\mathbf{V}^\top \mathbf{W}_s(v_0)\mathbf{V})^{-1}\mathbf{V}^\top \mathbf{W}_s(v_0)\mathbf{P} = \sum_{i=1}^{n} l_i(v_0; s)\, p_i \tag{12}$$

where $\mathbf{V}_0 = (1, v_0, v_0^2, \ldots, v_0^k)$ and the term $l_i(v_0; s)$ combines the local smoothing kernel $\mathcal{K}_s(v_0, \cdot)$ and the least squares operation for fitting the polynomial regression.

One can interpret s as the fraction of observations used in constructing the local fit at any point v_0. Small values of s will produce more local fits, while large values result in more global fits using all the observations and the resulting model will be similar to the polynomial regression. Fig. 3 shows the locally weighted linear and quadratic regression models of power on wind speed for observations generated by turbines V82 and FL-255. The cross-validation technique is used to obtain the optimum value of s for each data set. By comparing Fig. 2 with Fig. 3, we observe that locally weighted polynomial regression models reduce the bias of polynomial regression models, especially at the boundaries.

However, the bias reduction is obtained at a cost of variance increase. Therefore, for choosing k, one needs to pay careful attention to the bias–variance tradeoff of the fitted models [28], [29].

C. Spline Regression

Although locally weighted polynomial regression method could result in an appropriate approximation to the manufacturer power curve, in practice, one might not have enough control on the curvature of the fitted power curve to provide a desirable approximation to the nonlinear nature of the relationship between the generated power and the wind speed. One way to achieve more flexibility and provide more control on the curvature of the fitted power curve is to use piecewise polynomials. A piecewise polynomial regression involves fitting separate low-degree polynomials over different regions of the wind speeds. To this end, one needs to specify K different breakpoints, known as knots, throughout the range of the wind speeds and then fit $K+1$ different polynomial regression models. To fit a smooth and continuous piecewise degree-k polynomial regression, we need to put the constraints that the first $k-1$ derivatives of the fitted power curve to be continuous. This can be achieved by using the polynomial spline regression function which is defined as

$$p_i = \beta_0 + \sum_{r=1}^{k} \beta_r v_i^r + \sum_{j=1}^{K} \beta_{k+j}(v_i - \zeta_j)_+^k + \epsilon_i \tag{13}$$

where $k \geq 1$ is the order of spline, $\zeta_1, \ldots, \zeta_K$ are a set of prespecified knots, and the function $(\cdot)_+^k$ denotes a truncated power function as follows:

$$(v_i - \zeta_j)_+^k = \begin{cases} (v_i - \zeta_j)^k, & v_i > \zeta_j, \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

The most popular spline regression is the cubic spline corresponding to the choice of $k = 3$ in (13). Cubic spline regression models have been used for power curve modeling by [30] and are reliable to predict wind turbine power with smooth power curves. Note that for a sample of size n, the general spline regression model in (13) can be written as

$$\mathbf{P} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{15}$$

with $\mathbf{Z}_{n\times(K+k+1)} = (\mathbf{V}_{n\times(k+1)}, \mathbf{U}_{n\times K})$, where $\mathbf{U}_{n\times K}$ is a matrix with elements $(v_i - \zeta_j)_+^k$, $i = 1, \ldots, n$ and $j = 1, \ldots, K$. Hence, the model parameters can be estimated by the least squares method to obtain

$$\widehat{\boldsymbol{\beta}} = (\mathbf{Z}^\top \mathbf{Z})^{-1}\mathbf{Z}^\top \mathbf{P} \tag{16}$$

and the fitted power curve at speed v_i is $\widehat{f}(v_i) = \mathbf{Z}_i\widehat{\boldsymbol{\beta}}$, where $\mathbf{Z}_i = (1, v_i, \ldots, v_i^k, (v_i - \zeta_1)_+^k, \ldots, (v_i - \zeta_K)_+^k)$ is the ith row of the matrix $\mathbf{Z}$.

The normal equations associated with the truncated power basis are highly ill-conditioned resulting in inaccuracies in the calculation of $\widehat{\boldsymbol{\beta}}$. For computational purposes, we use the B-spline basis and reformulate (13) as $\mathbf{P} = \mathbb{B}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbb{B}$ is a matrix with its (i, j)th element being $B_j^k(v_i)$, where

$$B_j^k(v_i) = \frac{(v_i - \zeta_j)}{(\zeta_{j+k} - \zeta_j)} B_j^{k-1}(v_i) + \frac{(\zeta_{j+k+1} - v_i)}{(\zeta_{j+k+1} - \zeta_{j+1})} B_{j+1}^{k-1}(v_i)$$

for $j = -k, -k+1, \ldots, K$, $\zeta_0 = \zeta_{-1} = \cdots = \zeta_{-k} = \min\{v_i, i = 1, \ldots, n\}$, and $\zeta_K + 1 = \max\{v_i, i = 1, \ldots, n\}$. Also, $B_j^0(v_i)$ are the natural basis for piecewise constant functions. The least squares estimates of $\boldsymbol{\beta}$ is then given by

$$\widehat{\boldsymbol{\beta}} = (\mathbb{B}^\top \mathbb{B})^{-1}\mathbb{B}^\top \mathbf{P} \tag{17}$$

which is more feasible for computational purposes. For other formulations of the spline regression, refer [31]. As depicted

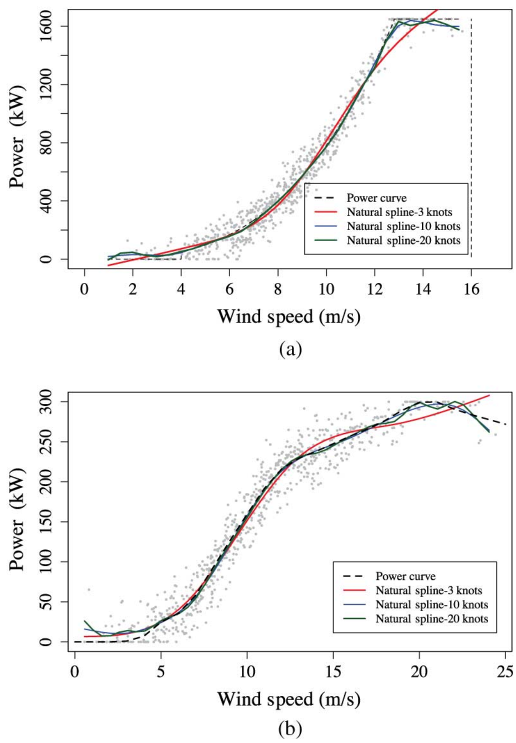


Fig. 4. Natural cubic spline power curve fitting with 3, 10, and 20 knots for the power data generated from turbines V82 (a) and FL-255 (b). The black-dashed curves are theoretical power curves from the manufacturers.

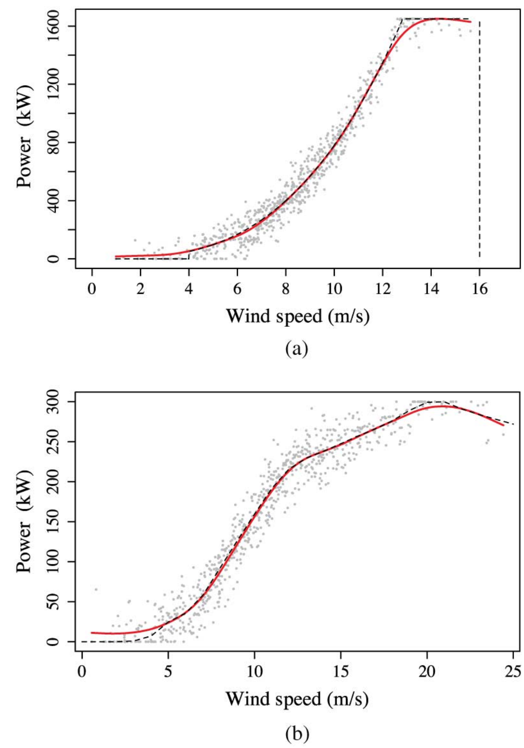


Fig. 5. Fitted power curves (solid red curve) for (a) turbine V82 and (b) turbine FL-255 using penalized smoothing spline with the theoretical power curves (back-dashed curve).

in Fig. 4, spline regression often leads to superior results over polynomial regression. This is because spline regression introduces flexibility by increasing the number of knots but keeping the degree of polynomial fixed. If the function $f(\cdot)$ changes rapidly in a region of v, one can add more knots to capture the change. However, spline regression tends to behave erratically beyond the boundary knots compared with the corresponding global polynomial regression in those regions [32]. Natural spline regressions, which are constrained to be linear beyond the boundary knots, provide a useful tool to overcome this problem. Fig. 4 shows the natural cubic spline power curve models fitted to the power data for turbines V82 and FL-255 when using 3, 10, and 20 equally spaced knots. We observe that, if the number and location of knots are chosen badly, spline regression will result in a poor fit. Several methods are proposed in the literature to obtain algorithms for optimizing over the number and location of knots, as shown by [33].

D. Penalized Spline Regression

To address the challenge of choosing the number and the location of knots in spline regression, we propose to use a penalized spline regression model for fitting wind turbine power curves. The idea is to use spline regression with a fixed basis dimension at a size slightly larger than it is necessary (e.g., fixed quantiles of wind variable), but to control the power curve smoothness by adding a penalty to the least squares fitting objective. In other words, we fit a power curve to the data points by minimizing

$$\frac{1}{n}\sum_{i=1}^{n}(p_i - f(v_i))^2 + \lambda\int\{f^{''}(t)\}^2\,dt \qquad (18)$$

where λ is a fixed smoothing parameter and $f^{''}(\cdot)$ is the second derivative of $f(\cdot)$. The first term in (18) measures the goodness of fit of the curve, while the second term measures the roughness of $f(\cdot)$. The smoothing parameter λ balances the tradeoff between goodness of fit and roughness of the curve. If the penalty is zero, we obtain a curve that interpolates the data points. If the penalty is infinite, we obtain an ordinary least squares fit to the data. The common way of choosing λ is by cross-validation. In many cases, the penalty term can be written as a quadratic form $\boldsymbol{\beta}^\top\mathbf{D}\boldsymbol{\beta}$, where $\mathbf{D}$ is a matrix of known coefficients. The estimated model parameters under penalized spline regression are given by

$$\hat{\boldsymbol{\beta}}(\lambda) = (\mathbf{Z}^\top\mathbf{Z} + \lambda\mathbf{D})^{-1}\mathbf{Z}^\top\mathbf{P} \qquad (19)$$

and the fitted power curve is

$$\hat{\mathbf{P}}(\lambda) = \mathbf{Z}\hat{\boldsymbol{\beta}}(\lambda) = \mathbf{Z}(\mathbf{Z}^\top\mathbf{Z} + \lambda\mathbf{D})^{-1}\mathbf{Z}^\top\mathbf{P} = \mathbf{H}(\lambda)\mathbf{P} \qquad (20)$$

where $\mathbf{H}(\lambda)$ is the hat matrix. In this paper, we choose the penalty function to be $\sum_{j=1}^{K}\beta_{p+j}^2$, which results in $\mathbf{D} = \mathrm{diag}(\mathbf{0}_{p+1}, \mathbf{1}_K)$ [34]. To obtain a suitable value of λ, we use the generalized cross-validation statistic

$$\mathrm{GCV}(\lambda) = \frac{\sum_{i=1}^{n}(p_i - \hat{p}_i(\lambda))^2}{n\,[1 - \lambda n^{-1}\mathrm{trace}\{\mathbf{H}(\lambda)\}]^2} \qquad (21)$$

where $\hat{p}_i(\lambda)$ is the ith element of $\hat{\mathbf{P}}(\lambda)$ in (20). Here, $\mathrm{trace}\{\mathbf{H}(\lambda)\}$ is called the effective degrees of freedom of the fit. We choose a suitable value of λ by computing $\mathrm{GCV}(\lambda)$ for a grid of λ values and choosing the minimizer over the grid. Fig. 5 depicts the power curves fitted to the simulated observations from turbines V82 and FL-255 using penalized smoothing spline

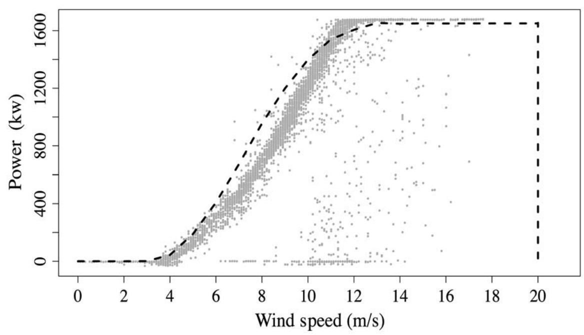


Fig. 6. Turbine T_1 10-min measured power data versus measured wind speed. The manufacturer power curve is shown in dashed line with characteristic speeds: cut-in ($V_c = 3.5$), rated ($V_r = 13$), and cut-out ($V_s = 20$).

when the smoothing parameter is obtained using generalized cross-validation technique. As shown in this figure, penalized smoothing spline provides the best fit, and the obtained power curves are similar to the manufacturers. In the proposed method, the difficulty of choosing the number and the location of knots in cubic spline regression is reduced to a simple problem and can be solved by cross-validation technique.

III. Real Data Application

In this section, we apply the methods in Section II on proprietary wind power data of a wind farm in North America. The wind power plant (WPP) includes over five dozen identical wind turbines with the rated capacity of 1.7 MW and the hub height of 80 m spread over an area of over 90 km^2. The cut-in speed, rated speed, and cut-out speed of the turbines are 3.5, 13, and 20 m/s, respectively. There are three meteorological (MET) towers located in the WPP collecting wind speeds, wind directions, air density, and humidity at 10-min average. We have selected four wind turbines $(T_1, \ldots, T_4)$ of the WPP to analyze the performance of the presented methods. To more accurately represent the wind data, turbines T_1, T_2, and T_3 are chosen near the MET towers, while the turbine T_4 is away from the towers. We use two raw data sets of 4320 pairs representing 1 month 10-min averaged data in June–July 2006, and 1 month in December–January 2007. This offers a more realistic representation of the performance of the turbines.

Fig. 6 shows that the measured data of wind power versus wind speed do not exactly match the power generation curve provided by the manufacturer. This can be explained by the difference between the standard test conditions and the conditions of the actual site. Wind direction, vertical wind speed profile, horizontal uniformity of the wind speed across the face of the turbine, as well as the distance between the MET tower and the wind turbine, maintenance operations, and aging components are among other influencing factors [35]. Similar to [36], we observe different types of data points in our raw data set that can be classified as per Table I. Data processing is required to filter the invalid data points. According to Table I, data type 1 is the desired data point. Types 4 and 5 of the data, where negative values for wind and power are observed, are filtered. To filter out the data types 2 and 3, for each wind speed datapoint, only the power values that lie within three standard deviations of the average power value at that speed are included.

TABLE I
Wind Power and Speed Raw Data Classification

Type	Data description
1	Data points following the pattern of turbine's power curve.
2	Data points with high wind speed and low power values.
3	Data points with low wind speed and high power values.
4	Data points with negative wind speed values.
5	Data points with negative wind power values.

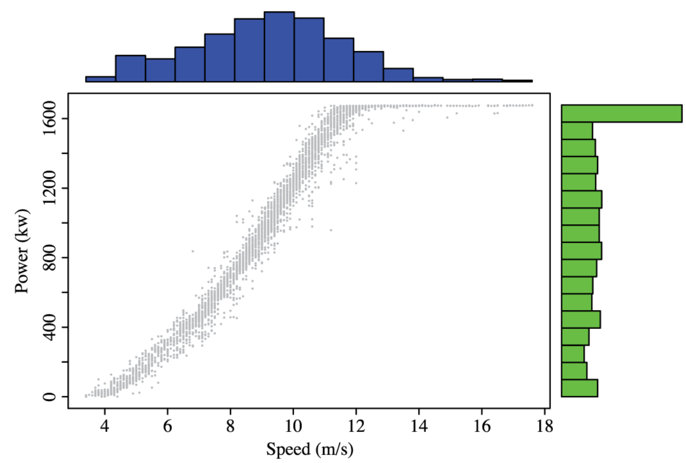


Fig. 7. The scatter plot of the wind speed and the generated power with the marginal histograms associated with each variable by turbine T_1 for 1 month in winter.

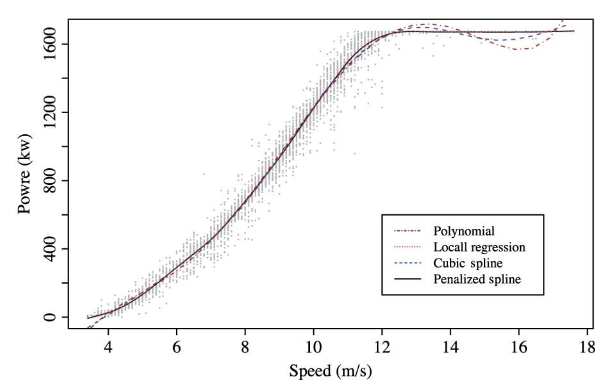


Fig. 8. The scatter plot of the wind speed and the generated power by turbine T_1 for 1 month in winter with the fitted power curves using proposed methods.

Fig. 7 shows the obtained scatter plot of 3663 data points from turbine T_1 for the month in winter after performing the filtering procedure. This figure also depicts the marginal histograms associated with the wind speed distribution on the horizontal axis and the generated power distribution on the vertical axis. Fig. 8 shows the four models proposed in this paper applied on the filtered wind speed and the generated wind power data.

There are several statistical metrics that can be used as appropriate measures of performance for the fitted power curves such as the root-mean-squared error (RMSE), normalized mean absolute percentage error (NMAPE), symmetric mean absolute percentage error (sMAPE), the mean absolute

TABLE II
RESULTS FOR SUMMER AND WINTER 10-MIN AND HOURLY AVERAGED DATA FOR TURBINES T_1, T_2, T_3 , AND T_4 FOR THE METHODS POLYNOMIAL REGRESSION (PR), LOCALLY WEIGHTED REGRESSION (LR), CUBIC SPLINE (CS), AND PENALIZED SPLINE (PS)

Time		Method	Turbine T_1		Turbine T_2		Turbine T_3		Turbine T_4		Final
			NMAPE	RMSE	NMAPE	RMSE	NMAPE	RMSE	NMAPE	RMSE	rank
Hourly	Summer	PR	1.805 (4)	48.69 (4)	1.777 (4)	47.86 (4)	1.715 (4)	48.69 (4)	2.898 (1)	46.58 (3)	4
		LR	1.647 (2)	47.79 (1)	1.612 (2)	45.11 (3)	1.680 (2)	47.70 (1)	2.968 (4)	38.34 (1)	2
		CS	1.668 (3)	47.91 (3)	1.619 (3)	45.10 (2)	1.689 (3)	47.91 (3)	2.910 (2)	38.67 (2)	3
		PS	1.626 (1)	47.81 (2)	1.595 (1)	44.89 (1)	1.678 (1)	47.81 (2)	2.914 (3)	38.34 (1)	1
	Winter	PR	3.723 (4)	99.88 (4)	2.226 (3)	60.45 (4)	2.789 (4)	104.21 (4)	3.539 (4)	43.44 (4)	4
		LR	3.408 (1)	96.24 (2)	2.016 (2)	57.12 (1)	2.681 (2)	102.89 (2)	3.441 (1)	36.29 (2)	2
		CS	3.430 (3)	98.96 (3)	2.016 (2)	57.46 (3)	2.786 (3)	103.47 (3)	3.514 (3)	36.63 (3)	3
		PS	3.424 (2)	95.93 (1)	1.960 (1)	54.38 (2)	2.386 (1)	102.26 (1)	3.471 (2)	36.28 (1)	1
10–minute	Summer	PR	1.922 (4)	46.58 (3)	1.748 (4)	43.44 (4)	1.896 (4)	45.93 (4)	2.461 (4)	64.69 (4)	4
		LR	1.651 (2)	38.34 (1)	1.484 (2)	36.29 (2)	1.689 (2)	42.99 (2)	2.356 (3)	63.51 (2)	2
		CS	1.673 (3)	38.67 (2)	1.497 (3)	36.63 (3)	1.693 (3)	43.04 (3)	2.352 (2)	63.62 (3)	3
		PS	1.648 (1)	38.34 (1)	1.480 (1)	36.28 (1)	1.682 (1)	42.89 (1)	2.349 (1)	63.38 (1)	1
	Winter	PR	2.858 (3)	69.23 (4)	2.217 (4)	49.05 (4)	2.575 (4)	63.23 (1)	3.398 (4)	93.40 (4)	4
		LR	2.555 (1)	67.74 (3)	1.682 (1)	41.93 (2)	2.281 (1)	65.74 (3)	3.079 (2)	89.25 (1)	2
		CS	2.705 (2)	67.37 (1)	2.019 (3)	45.04 (3)	2.396 (3)	67.37 (4)	3.389 (3)	92.37 (3)	3
		PS	2.555 (1)	65.62 (2)	1.683 (2)	41.87 (1)	2.301 (2)	65.62 (2)	3.071 (1)	90.22 (2)	1

error (MAE), and the coefficient of determination (R^2) [18], [37]. In this paper, we use RMSE and NMAPE given by

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - \widehat{p}_i)^2} \quad (22)$$

and

$$\text{NMAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|p_i - \widehat{p}_i|}{\max_{j=1}^{n}\{\widehat{p}_j\}} \times 100 \quad (23)$$

where p_i is the observed wind power and $\widehat{p}_i$ is the estimated value of the power using the underlying method. By performing the error analysis, the values of RMSE and NMAPE for the polynomial regression (PR), locally weighted regression (LR), cubic spline (CS), and penalized spline (PS) methods are presented here. In Table II, we show the results of the analysis for two data sets representing 10-min and hourly averaged data in June–July 2006 as well as December–January 2007. For the hourly data, the same data sets are used and are averaged hourly to investigate the influence of data resolutions on the error measures. Table II also ranks the performance of all four methods based on the calculated measures, where the smaller values are desirable. The ranking numbers are shown in the parenthesis next to RMSE and NMAPE values. We also provide the overall performance ranking of each method. As shown in the table, penalized spline regression outperforms all other methods addressed in this study. Table II also suggests that locally weighed polynomial regression is dominating polynomial regression and the cubic spline. Comparing the values of RMSE and NMAPE for summer and winter shows the error measures are greater in the winter data for the same generator. This can be explained by the impact of weather condition and cold temperature on the atmospheric parameters and on the mechanical efficiency of the machine.

The overall performance of the error metrics for turbine T_4, which is the one located away from the MET towers, is similarly following the performance ranking in Table II. However, in the winter data set, the RMSE and NMAPE values in hourly data are noticeably higher than those of 10-min data. By repeating the analysis with other statistical metrics such as MAE, sMAPE, and R^2 (results are not presented here), we obtained similar patterns in the rankings of the presented methods.

IV. CONCLUSION

Accurate modeling of the wind turbine power curves is an important tool in the wind energy industry that can be used for assessment and monitoring of the turbine's performance, power forecasting, as well as sizing the storage capacity for wind power integration. We have presented parametric and nonparametric regression models for estimating wind turbine power curves. Polynomial regression is used as the benchmark parametric method for power curve fitting. We have shown that polynomial regression models are limited by their global nature and are very sensitive to outliers. Also, finding a good fit to the empirical data requires a high-degree polynomial regression model which can cause an overfitting to the observed data. Locally weighted polynomial regression is introduced to address the issues of the global nature in polynomial regression and its sensitivity to outliers within the observations. Spline regression method, which is based on piecewise polynomial regression models, is then examined to achieve more flexibility for fitting the power curve. In this method, an important issue is finding the number and the location of knots to provide a good fit to the empirical power curve. To this end, we propose a penalized spline regression model that reduces these problems to a simple problem of choosing a single parameter that can be determined using a cross-validation technique. The performance of the proposed methods is evaluated based on two simulated random data sets with normal errors as well as an operational wind power data for a wind farm in North America. Four wind turbines are selected to analyze the performance of the presented methods, of which, three turbines are chosen near the MET towers, while the last one is located away from the towers. The accuracy of each method is evaluated using the RMSE and NMAPE metrics. The results of this study suggest that penalized spline regression presents a better performance over the other analyzed methods. The outcome of this study can be used in various applications such as turbine performance monitoring, power forecasting, and sizing the storage capacity for wind power integration. Currently, this method is being applied to the battery sizing model in our research group at the University of Manitoba, Winnipeg, MB, Canada [7].

REFERENCES

[1] International Energy Agency (IEA), "World energy outlook—2013," IEA, Paris, Tech. Rep., 2013.

[2] International Energy Agency (IEA), "Technology roadmap—Wind energy," IEA, Paris, Tech. Rep., 2013.

[3] A. Ipakchi and F. Albuyeh, "Grid of the future," *IEEE Power Energy Mag.*, vol. 7, no. 2, pp. 52–62, Mar. 2009.

[4] A. Kusiak, A. Verma, and X. Wei, "Wind turbine capacity frontier from scada," *Wind Syst. Mag.*, vol. 3, no. 9, pp. 36–39, 2012.

[5] A. Benatiallah, L. Kadia, and B. Dakyob, "Modelling and optimisation of wind energy systems," *Jordan J Mech. Ind. Eng.*, vol. 4, no. 1, 2010.

[6] Z. Zhang, Q. Zhou, and A. Kusiak, "Optimization of wind power and its variability with a computational intelligence approach," *IEEE Trans. Sustain. Energy*, vol. 5, no. 1, pp. 228–236, Jan. 2014.

[7] S. Shokrzadeh, M. Jafari-Jozani, E. Bibeau, and T. Molinski, "An energy storage sizing algorithm for base-load power generation of wind farms," submitted for publication.

[8] R. Gasch and J. Twele, *Wind Power Plants: Fundamentals, Design, Construction and Operation*, 2nd ed. Berlin, Germany: Springer-Verlag, 2012.

[9] J. F. Manwell, J. G. McGowan, and A. L. Rogers, *Wind Energy Explained: Theory, Design and Application*, 2nd ed. Hoboken, NJ, USA: Wiley, 2009.

[10] C. Carrillo, A. O. Montao, J. Cidrs, and E. Daz-Dorado, "Review of power curve modelling for wind turbines," *Renew. Sustain. Energy Rev.*, vol. 21, pp. 572–581, 2013.

[11] M. Lydia, S. S. Kumar, A. I. Selvakumar, and G. E. P. Kumar, "A comprehensive review on wind turbine power curve modeling techniques," *Renew. Sustain. Energy Rev.*, vol. 30, pp. 452–460, 2014.

[12] A. Kusiak, H. Zheng, and Z. Song, "On-line monitoring of power curves," *Renew. Energy*, vol. 34, no. 6, pp. 1487–1493, 2009.

[13] M. Hollander, D. Wolfe, and E. Chicken, *Nonparametric Statistical Methods*, 3rd ed. Hoboken, NJ, USA: Wiley, 2014.

[14] M. G. Khalfallah and A. M. Koliub, "Suggestions for improving wind turbines power curves," *Desalination*, vol. 209, no. 3, pp. 221–229, 2007.

[15] P. Giorsetto and K. F. Utsurogi, "Development of a new procedure for reliability modeling of wind turbine generators," *IEEE Trans. Power App. Syst.*, vol. PAS-102, no. 1, pp. 134–143, Jan. 1983.

[16] R. Chedid, H. Akiki, and S. Rahman, "A decision support technique for the design of hybrid solar-wind power systems," *IEEE Trans. Energy Convers.*, vol. 13, no. 1, pp. 76–83, Mar. 1998.

[17] M. Lydia, A. Selvakumar, S. Kumar, and G. Kumar, "Advanced algorithms for wind turbine power curve modeling," *IEEE Trans. Sustain. Energy*, vol. 4, no. 3, pp. 827–835, Jul. 2013.

[18] A. Marvuglia and A. Messineo, "Monitoring of wind farms' power curves using machine learning techniques," *Appl. Energy*, vol. 98, no. 0, pp. 574–583, 2012.

[19] T. Üstüntaş and A. D. Şahin, "Wind turbine power curve estimation based on cluster center fuzzy logic modeling," *J Wind Eng. Ind. Aerodyn.*, vol. 96, no. 5, pp. 611–620, 2008.

[20] A. Kusiak, H. Zheng, and Z. Song, "Models for monitoring wind farm power," *Renew. Energy*, vol. 34, no. 3, pp. 583–590, 2009.

[21] T. J. Hastie and R. J. Tibshirani, *Generalized Additive Models*. London, U.K.: Chapman & Hall, 1990.

[22] B. Ai, H. Yang, H. Shen, and X. Liao, "Computer-aided design of pv/wind hybrid system," *Renew. Energy*, vol. 28, no. 10, pp. 1491–1512, 2003.

[23] S. Diaf, D. Diaf, M. Belhamel, M. Haddadi, and A. Louche, "A methodology for optimal sizing of autonomous hybrid pv/wind system," *Energy Policy*, vol. 35, no. 11, pp. 5708–5718, 2007.

[24] J. Walker and N. Jenkins, *Wind Energy Technology*. Hoboken, NJ, USA: Wiley, 1997.

[25] T. Ackermann and L. Sder, "Wind energy technology and current status: A review," *Renew. Sustain. Energy Rev.*, vol. 4, no. 4, pp. 315–374, 2000.

[26] T. Ackermann, *Wind Power in Power Systems*, 2nd ed. Hoboken, NJ, USA: Wiley, 2012.

[27] A. Betz, "Das maximum der theoretisch moglichen ausnutzung des windes durch windmotoren, the maximum of the theoretically possible exploitation of the wind by wind motors," *Z. Gesamte Turbinenwesen*, vol. 26, pp. 307–309, 1920.

[28] W. S. Cleveland and S. J. Devlin, "Locally weighted regression: An approach to regression analysis by local fitting," *J Amer. Stat. Assoc.*, vol. 83, pp. 596–610, 1988.

[29] W. S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *J Amer. Stat. Assoc.*, vol. 74, pp. 829–836, 1979.

[30] V. Thapar, G. Agnihotri, and V. K. Sethi, "Critical analysis of methods for mathematical modelling of wind turbines," *Renew. Energy*, vol. 36, no. 11, pp. 3166–3177, 2011.

[31] G. Wahba, *Spline Models for Observational Data*. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 1990.

[32] T. J. Hastie, R. J. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, vol. 27, New York, NY, USA: Springer, 2005, pp. 83–85.

[33] J. H. Friedman and B. W. Silverman, "Flexible parsimonious smoothing and additive modeling," *Technometrics*, vol. 31, no. 1, pp. 3–21, 1989.

[34] D. Ruppert, "Selecting the number of knots for penalized splines," *J Comput. Graph. Statist.*, vol. 11, no. 4, pp. 735–757, 2002.

[35] S. Li, D. Wunsch, E. O'Hair, and M. Giesselmann, "Using neural networks to estimate wind turbine power generation," *IEEE Trans. Energy Convers.*, vol. 16, no. 3, pp. 276–282, Sep. 2001.

[36] Z. Liu, W. Gao, Y.-H. Wan, and E. Muljadi, "Wind power plant prediction by using neural networks," in *2012 IEEE Energy Convers. Congr. Expo. (ECCE)*, 2012, pp. 3154–3160.

[37] A. Kusiak and X. Wei, "A data-driven model for maximization of methane production in a wastewater treatment plant." *Water Sci. Technol.*, vol. 65, no. 6, 2012.

Shahab Shokrzadeh (S'09) received the B.Sc. degree in mechanical engineering from Iran University of Science and Technology, Tehran, Iran, in 2004, and the M.Sc. degree in production engineering and management from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 2009. He is currently pursuing the Ph.D. degree in mechanical engineering at the University of Manitoba, Winnipeg, MB, Canada.

His research interests include the areas of renewable energy technologies and sustainable transportation, and is developing expertise in applications of parametric and nonparametric statistical methods in the wind energy industry.

Mohammad Jafari Jozani received the B.Sc. degree in statistics from Allameh Tabatabaie University, Tehran, Iran, and the M.Sc. and Ph.D. degrees in mathematical statistics from Shahid Beheshti University, Tehran, Iran, in 1999, 2001, and 2006, respectively.

Currently, he is an Associate Professor with the Department of Statistics, University of Manitoba, Winnipeg, MB, Canada. His research interests include Bayesian analysis as well as parametric and nonparametric inference using complex sampling designs and recently developed interests in the application of modern statistical methods in renewable energy and stochastic modeling of wind turbines power forecasting and performance monitoring.

Eric Bibeau received the B.Sc. degree in mechanical engineering from McGill University, Montreal, QC, Canada, in 1986, and the M.Sc. and Ph.D. degrees in mechanical engineering from the University of British Columbia, Vancouver, BC, Canada, in 1988 and 1993, respectively.

Currently, he is an Associate Professor with the Department of Mechanical Engineering, University of Manitoba, Winnipeg, MB, Canada, and holds the NSERC/Manitoba Hydro Chair in Alternative Energy. He specializes in developing power systems using biomass feedstocks, industrial waste heat, and kinetic energy from rivers for distributed energy applications, in developing blade deicing mitigation strategies for wind turbines, and in developing expertise in electric vehicles and district energy systems as a means to increase the renewable energy ratio in Canada to simultaneously address climate change and peak oil energy drivers.

Dr. Bibeau is a Registered Professional Engineer in the Province of Manitoba.