Yaw-Adjusted Wind Power Curve Modeling: A Local Regression Approach

Praanjal Nasery[a], Ahmed Aziz Ezzat[b]

[a] *The Electric Power Research Institute, Palo Alto, 94304, CA, USA*
[b] *Department of Industrial & Systems Engineering, Rutgers University, Piscataway, 08854, NJ, USA*

Abstract

Accurate estimation of wind power curves using field data is instrumental to several wind farm operations including productivity assessment, power output estimation, operations and maintenance, among others. Existing methods for estimating wind power curves mainly rely on environmental variables (e.g., wind speed, direction, density) as inputs to construct the wind-to-power relationship. This paper attempts to integrate yaw misalignment as an additional input to power curve models, constructing what is referred to hereinafter as "yaw-adjusted wind power curves." Our analysis shows that integrating yaw misalignment into power curves is non-trivial, largely due to the overwhelming impact of environmental variables (mainly wind speed) on a turbine's power output, which obscures the secondary effect of yaw errors on power production. In response, we propose a local-regression-based method which reconstructs the yaw-to-power relationship *conditional* on an effective neighborhood of environmental variables. Tested on operational data from two onshore wind turbines in France, our proposed approach achieves significant improvements, in terms of power estimation accuracy, relative to a set of prevalent statistical- and machine-learning-based power curve models.

Keywords: Local Regression, Power Curve, Wind Energy, Yaw Error

1. Introduction

Wind power continues to be one of the fastest growing sources of clean energy worldwide. Despite its rapid growth, the uncertainty associated with generating electricity from wind remains a fundamental barrier that impedes its large-scale penetration into modern-day power systems. This uncertainty

stems from two main sources: (i) uncertainty in the wind resource, which is, arguably, one of the most challenging meteorological processes to predict, especially at the turbine level [1, 2]; and (ii) uncertainty in the wind-to-power conversion process, i.e. the conversion of the hub-height wind conditions experienced by a wind turbine into power output [3, 4, 5]. This work focuses on the latter, i.e., to develop accurate models for turbine-specific power output estimation. In the wind energy industry, an accurate estimation of a turbine's power production is important to several key wind farm operations, including performance assessment [6, 7], power output prediction [8, 9], asset monitoring and prognostics [10, 11], and maintenance scheduling [12, 13].

Theoretically, the wind power, P, produced by a wind turbine is expressed as in (1), where ρ denotes the air density, A_w is the rotor swept area, v is the hub-height wind speed, and C_p is the power coefficient, which, in turn, depends on the tip speed ratio λ and the blade pitch angle β.

$$P = \frac{1}{2}\rho A_w C_p(\lambda, \beta) v^3. \tag{1}$$

From (1), it is clear that the power production of a wind turbine is primarily dependent on wind speed, as evident by the cubic speed-to-power relationship, but also on other environmental variables such as air density. Moreover, (1) assumes that the turbine's rotor plane is perfectly perpendicular to the wind flow. In reality, this perfect alignment seldom happens, and the power produced by a wind turbine is further dependent on the yaw error (or misalignment), denoted hereinafter by γ [14].

In practice, the actual wind-to-power conversion process, however, does not fully adhere to the physical relationship in (1), but instead, follows the so-called "power curve", which is typically defined as the functional mapping relating the hub-height wind speed with the associated wind power, and comprises four main regions, as shown in Figure 1(a). In the first region, the turbine barely produces any power, as the wind speed is less than the so-called cut-in speed, $v_{cut\text{-}in}$. In modern-day wind turbines, cut-in speeds range between 3 m/s to 5 m/s. In the second region, the power output rapidly increases, first in a convex-shaped pattern, which then turns into a concave-shaped pattern. The wind speed in this region ranges between $v_{cut\text{-}in}$ and the rated speed, denoted in Figure 1(a) by v_{rated}, which refers to the minimum wind speed at which the maximal turbine capacity, P_{rated}, is reached. This second region is further divided into two sub-regions: Regions II-a (Maximum power point tracking or MPPT control) and II-b (Fixed speed control). In

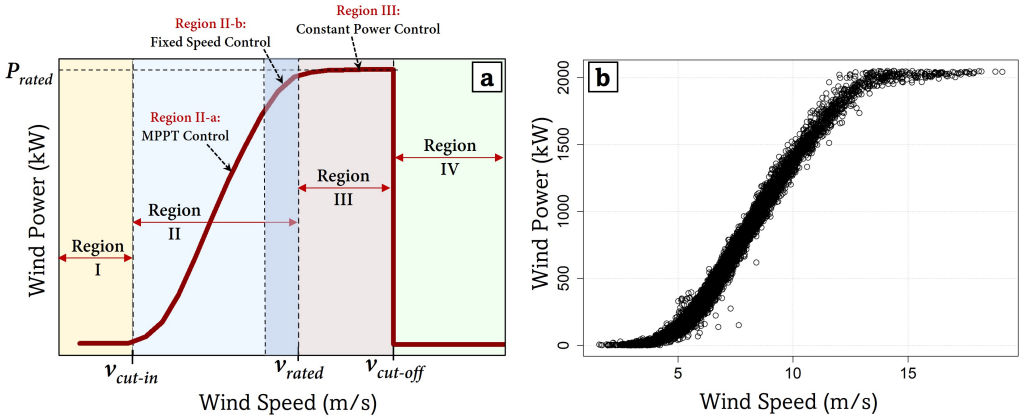

Figure 1: (a) Power curve of a typical wind turbine, with four regions defined by $v_{\text{cut-in}}$, v_{rated}, and $v_{\text{cut-off}}$.; (b) Actual wind speed versus power for a typical wind turbine.

the former, both the power output and the rotor speed are within their permissible operational limits, while in the latter, the rotor speed operates at its maximum limit, until the turbine reaches its rated power, P_{rated}. In the third region, the power produced is maintained at the rated power level for damage-mitigating purposes. This rated power is maintained until the cut-off speed, $v_{cut\text{-}off}$, is reached, after which the turbine shuts down to protect its components and structures from extreme wind conditions (Region IV).

In practice, estimating a turbine's power curve using field data is challenging, because the actual power curve is rarely deterministic as the one displayed in Figure 1(a). Instead, as shown in Figure 1(b), it is contaminated by large uncertainties which are attributed to multiple sources that can be broadly categorized into *environmental* and *operational* factors. Environmental factors refer to the impact of additional weather variables (other than wind speed) that may impact a turbine's power output, such as air density, wind direction, turbulence, among others. Operational factors refer to the parameters related to the response of the turbine to its external environment, such as yaw errors, wake effects, aging and degradation, and measurement noise. An accurate power curve estimation should therefore seek to construct a functional relationship which accurately links the power output of a turbine to a set of environmental and operational factors, while taking into account the observational noise [4].

Towards that end, there is an extensive literature on estimating power curves using field data. Broadly, those efforts can be classified into paramet-

ric versus nonparametric approaches [4, 15]. Parametric power curve models assume a pre-determined functional form with a fixed set of parameters. Prior efforts in parametric power curve modeling include but are not limited to piecewise linear models [16], polynomial regression models [3], and logistic functions [7]. Nonparametric models, on the other hand, impose little assumptions about the functional form of the input-output relationship. Nonparametric power curve modeling approaches include copula-based methods [17], cubic splines [18], artificial neural networks [19], Gaussian Processes [11], and k-nearest neighbors (k-NN) [7].

To date, the large majority of the literature, whether using parametric or nonparametric methods, primarily uses wind speed as the sole input to wind power curve models [20]. The main limitation of those univariate models is that they largely overlook the impact of other environmental variables apart from wind speed. Recent efforts have investigated the potential of integrating additional environmental variables into power curve models [21, 22, 23, 24], such as air density, direction, and turbulence intensity. Those efforts concluded that multivariate power curve models, i.e. those that consider the impact of several environmental variables, are superior, in terms of estimation accuracy, to those that rely on wind speed as the sole input.

Little research, however, has been conducted to explore the merit of integrating operational variables into power curve modeling. Barring few works which studied the integration of rotor speed and pitch angle in power curve estimation [25, 26], our work, up to our knowledge, constitutes the first attempt to formally integrate yaw misalignment in estimating wind power curves, yielding what we refer to hereinafter as "yaw-adjusted power curves."

We summarize the main contributions of this work as follows:

- We investigate the impact of yaw errors on power using field data. We discover that, counter-intuitively, a direct integration of yaw misalignment as an additional regressor into a multivariate power curve model does *not* improve the power curve estimation accuracy. We postulate that this is mainly due to the overwhelming impact of the environmental conditions (mainly wind speed) on a turbine's power output, which obscures the secondary effect of yaw errors on a the power output, and prevents classical statistical models from leveraging such finer-scale yaw-to-power correlations for improved power output estimation.

- Motivated by this observation, we propose a power curve estimation method based on local regression which reconstructs the yaw-to-power

relationship *conditional* on an effective neighborhood of environmental variables. We show that this approach in integrating yaw misalignment into wind power curves unearths its underlying impact on power generation, therefore enabling the power curve model to leverage it as a significant predictor of wind power production.

- We perform a series of extensive tests using real-world data from a wind farm in France to demonstrate the merit of our proposed approach relative to prevalent statistical- and machine-learning-based power curve models, eventually concluding that significant improvements in power estimation accuracy, ranging from 3.21% to 32.13% are realizable.

2. Data Description and Analysis

The data used herein comprises 10-min measurements collected in 2013 from the La Haute Borne-Vaudeville-le-Haut wind farm in north eastern France. The wind farm comprises four wind turbines, as shown in Figure 2, and is operated by Engie Green (a subsidiary of Engie), which provides open access to its data [27]. The turbines have a rated capacity of 2050 kW, 80-m hub height, and rotor diameter of 82 m. Of interest to us are the following SCADA variables: wind speed v (m/s), wind power P (kW), wind direction θ (°), nacelle angle ϕ (°), and air temperature T (K). Note that the wind direction and nacelle angle measurements have been pre-corrected in the dataset (we used the corrected versions of those variables) since SCADA measurements collected behind the turbine's rotor can be fairly noisy [28]. We also obtain co-located hourly air pressure measurements from Meteoblue [29], which we linearly interpolate to 10-min resolution. We filter the combined 10-min dataset by removing outliers (e.g., when P is negative or smaller than a certain threshold for a given wind speed).

The focus of our analysis is on two wind turbines (Turbine ID#: R80711 and R80721) for the period of September to December 2013. The turbines and time coverage were selected so that the turbines under study are upstream relative to the prevailing wind, which is found to be dominantly south westerly during this time of the year—See Figure 2.

We also compute two additional environmental variables: air density, denoted by ρ and determined as in (2), and turbulence intensity, denoted by I and computed as the ratio of the 10-min standard deviation of the wind

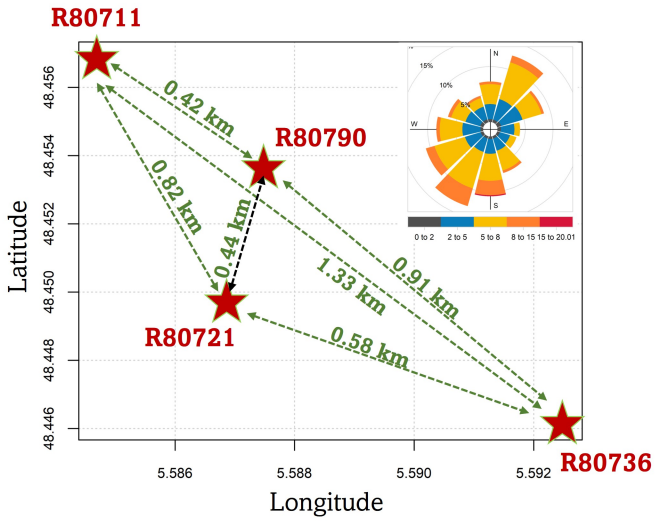

Figure 2: Map of the La Haute Borne-Vaudevulle-le-Haut wind farm operated by Engie. Turbine locations are depicted by red stars. Numbers on arrows denote inter-turbine distances. The top right inlet shows the wind rose plot at R80711, indicating a dominant south westerly wind during that time of the year.

speed (denoted by σ_v) to its 10-min average v, such that $I = \frac{\sigma_v}{v}$.

$$\rho = 1.225 \times \frac{288.15}{T} \times \frac{B}{1013.3}, \tag{2}$$

where B denotes the interpolated 10-min air pressure in mbar.

In constructing power curves, the wind energy industry utilizes an air density correction, wherein the raw 10-min wind speed v is transformed to its air-density-corrected counterpart $\tilde{v}$, through the following expression:

$$\tilde{v} = v \times \left(\frac{\rho}{1.225}\right)^{\frac{1}{3}}. \tag{3}$$

The yaw misalignment, γ, is computed as the absolute difference between the corrected wind direction θ, and the corrected nacelle angle ϕ. The large majority of data points (96.16%) are found to have yaw errors in the $[0°, 20°]$ interval, as shown in Figure 3a. Figure 3b shows a scatter plot of wind speeds, yaw errors, and power measurements, color-coded with the yaw misalignment values (only for $\gamma <= 20°$). Looking at Figure 3, there does not seem to

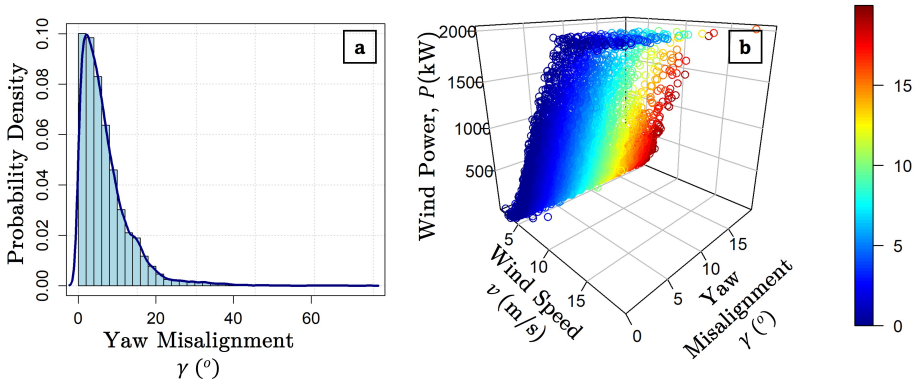

Figure 3: (a): Histogram of yaw misalignment values for turbine R80711. (b): 3D scatter plot of wind speed, yaw misalignment, and power for turbine R80711, color-coded by the yaw misalignment values ($\gamma \leq 20°$ is considered).

be stark negative correlations between the yaw error and the power output (-0.24 on a Pearson's correlation scale).

Based on Figure 3, one may mistakenly rush to the conclusion that yaw errors do not have a notable impact on the power output and should perhaps be dropped from power curve modeling. We conjecture, however, that this mainly attributed to the overwhelming impact of environmental variables (largely wind speed) on wind power, which obscures the secondary effect of yaw error on power output. Literature suggests that the power gain (or loss) due to yaw misalignment is typically in the $\sim$0-3% range and is a function of wind speed [30]. This means that slight changes in wind speed can easily "mask" the power variations caused by yaw errors. For example, if both wind speed and yaw error change by 1%, then the variation in power output P due to the change in wind speed alone will be much larger than that due to yaw error. This confounding effect statistically obscures the underlying impact of yaw errors on the turbine's power output.

To reveal the impact of yaw errors on the power output, we have to "marginalize" the effect of wind speed on power. To do so, we partition the wind speed domain into narrow intervals of 0.1 m/s, and examine the yaw-to-power relationship within each of those wind speed intervals. Within each interval, the variation in wind speed is small, thereby allowing us to "zoom in" and explore potential yaw-to-power correlations. Figure 4(a-c) shows the results of this partitioning exercise for three selected wind speed intervals: $[4.00, 4.10]$ m/s, $[5.60, 5.70]$ m/s, and $[7.30, 7.40]$ m/s. Within each wind

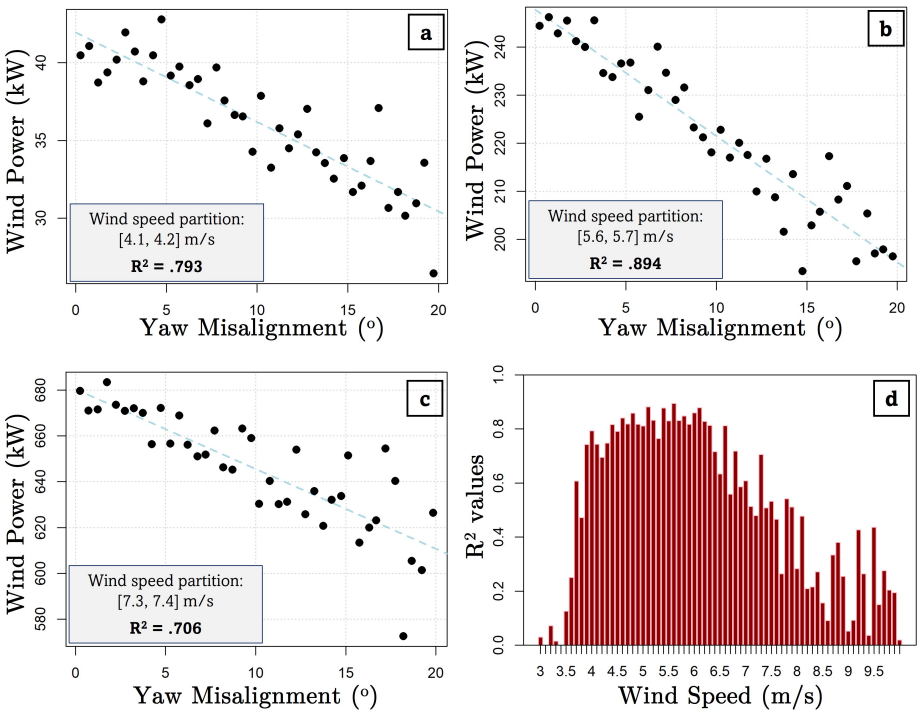

Figure 4: The yaw-to-power relationship for turbine R80711 within wind speed partitions of (a): $[4.00, 4.10]$ m/s, (b): $[5.60, 5.70]$ m/s, and (c): $[7.30, 7.40]$ m/s, showing a strong negative linear association (yaw errors binned into 0.5° intervals) (d): R^2 values of the yaw-to-power linear fits for all wind speed partitions between 3.00 and 10.00 m/s.

speed internval, the data suggest strong negative linear association between yaw errors and power output (with R^2 mostly in the 0.7-0.9 range). Figure 4d shows the R^2 values over all wind speed intervals considered.

The analysis suggests that yaw error has a notable impact on power output, particularly in the early and middle portions of Region II in a turbine's power curve. However, such impact is only noticeably revealed when conditioned on an effective neighborhood of environmental variables, or in other words, once the impact of other environmental variables (primarily wind speed) has been marginalized. A natural follow up question is how can we leverage such yaw-to-power correlations in estimating a turbine's power curve without relying on arbitrary partitions like those assumed in Figure 4. We address this question in Section 3 by proposing a neighborhood-based approach for yaw-adjusted power curve estimation.

3. A Local Regression Model for Yaw-adjusted Wind Power Curves

Our yaw-adjusted power curve model builds on a recently proposed multivariate power curve method called "the additive multivariate kernel" (AMK) [23]. This section therefore starts with a brief overview of the AMK method and then proceeds to show how we extend it using a local regression formulation in order to construct yaw-adjusted wind power curves.

3.1. Overview of The Additive Multivariate Kernel (AMK) Method

Let $\mathbf{x}^T = \{x_1, x_2, ..., x_q\}$ be a set of explanatory variables which will be used as inputs to the power curve model. For instance, $\mathbf{x}$ can contain wind speed, direction, air density, among others. Broadly speaking, our power curve estimation task is to learn a functional mapping $f : \mathbb{R}^q \rightarrow \mathbb{R}$ which relates $\mathbf{x}$ to the power output P, given a set of training data denoted by $\mathcal{D} = \{\mathbf{x}_i^T, P(\mathbf{x}_i)\}_{i=1}^n$. The estimated response surface, $\hat{f}$, constitutes the final power curve which will be used to make turbine-specific power predictions.

To integrate additional environmental variables beside wind speed (e.g., wind direction, air density) into wind power curves, kernel regression (KR) has been proposed in the past, due to its simplicity and attractive local modeling capabilities [31]. KR is essentially a nonparametric local regression model which uses the Nadaraya-Watson estimator [32], expressed in (4), to obtain a localized, yet smooth estimate of $f(\mathbf{x}^*)$ at a target input $\mathbf{x}^*$.

$$\hat{f}(\mathbf{x}^*) = \frac{\sum_{i=1}^n K(\mathbf{x}^*, \mathbf{x}_i) P(\mathbf{x}_i)}{\sum_{i=1}^n K(\mathbf{x}^*, \mathbf{x}_i)}, \tag{4}$$

where $K(\mathbf{x}^*, \mathbf{x}_i)$ is a kernel function, which, in essence, is a measure of similarity between the target input $\mathbf{x}^*$ and the ith training data point $\mathbf{x}_i$, thereby dictating the contribution of the latter in determining $\hat{f}(\mathbf{x}^*)$. A popular choice for $K(\mathbf{x}^*, \mathbf{x}_i)$ is the multivariate Gaussian kernel defined as in (5).

$$K(\mathbf{x}^*, \mathbf{x}_i) = \prod_{j=1}^q K(x_j^*, x_j) = \prod_{j=1}^q \exp\left(-\frac{1}{2} \frac{||x_{ij} - x_j^*||}{\lambda_j^2} \right), \tag{5}$$

where x_{ij} is the ith observation of the variable x_j, while x_j^* is the value of the jth variable in $\mathbf{x}^*$. The set of parameters, $\lambda_1, ..., \lambda_q$ dictate the width of the univariate kernels, or in other words, the size of the effective local neighborhood, whereas $|| \cdot ||$ denotes the Euclidean distance.

KR has been successfully used to model bivariate power curves, with wind speed and direction as inputs [31]. If more variables beyond speed and direction are introduced, the multiplicative form of the kernel in (5) can cause the KR method to run into the so-called "curse of dimensionality," wherein the number of data points within each high-dimensional neighborhood does not allow for a reliable (or in some cases, even feasible) estimation of $\hat{f}(\mathbf{x}^*)$. To remedy this limitation, the Additive Multivariate Kernel (AMK) method was proposed by Lee *et al.* [23] to integrate several environmental variables (in their paper, up to seven variables) like wind speed, direction, air density into a multivariate power curve model.

Instead of relying on the multiplicative form of (4), AMK proposes to compute $\hat{f}(\mathbf{x}^*)$ as in (6).

$$\hat{f}(\mathbf{x}^*) = \frac{1}{q-2}\left[\hat{f}_1(x_1^*, x_2^*, x_3^*) + ... + \hat{f}_{q-2}(x_1^*, x_2^*, x_q^*)\right], \tag{6}$$

wherein $\hat{f}_1(x_1^*, x_2^*, x_3^*)$, ..., $\hat{f}_{q-2}(x_1^*, x_2^*, x_q^*)$ are defined as in (4) using three-dimensional Gaussian kernels. AMK chooses to always fix the two variables x_1 and x_2 in each of the multiplicative kernels as the wind speed and direction respectively, such that $x_1 = \tilde{v}$ and $x_2 = \theta$, while x_3, ..., x_q denote additional environmental variables such as density, humidity, or turbulence intensity. Using the formulation in (6), AMK still uses multiplicative kernels to capture interactions in the input space, but restricts them to be the product of three univariate kernels in order to ensure scalability in higher dimensions. As a result, AMK was shown to yield significant improvements in estimating turbine-specific power relative to univariate and bivariate power curve models. Readers can refer to [23, 33] for more details of the AMK method.

3.2. Towards Yaw-adjusted Power Curves: The YAMK Method

Our method, the *Yaw-adjusted Additive Multivariate Kernel*, or in short YAMK, extends the AMK modeling framework to effectively integrate yaw misalignment into a multivariate wind power curve. The idea behind YAMK is to construct localized yaw-to-power regression models within the local neighborhoods defined by AMK. This formulation is inspired by the preliminary analysis in Section 2 wherein the effect of γ on P is only unearthed once the impact of other environmental conditions has been marginalized.

In our work, we let $\mathbf{x} := \{\tilde{v}, \theta, \rho, I\}$, that is, the set of all environmental variables. YAMK starts with a similar formulation to that of AMK expressed

in (6). The fundamental difference between YAMK and AMK, however, lies in how f_1, ..., f_q are defined. Instead of relying on locally weighted averages as in (4), YAMK defines them as locally weighted polynomial regression models which take the yaw misalignment, γ, as a direct input. That is, $\hat{f}_j(\mathbf{x}^*)$, $\forall$ $j = 1, ..., q-2$ is defined as in (7).

$$\hat{f}_j(\mathbf{x}^*) = g(\mathbf{x}^*)^T(\mathbf{G}^T\mathbf{K}\mathbf{G})^{-1}\mathbf{G}^T\mathbf{K}\mathbf{P}, \tag{7}$$

where $\mathbf{P}$ is the $n\times 1$ vector of power outputs, defined as $\mathbf{P} = \{P(\mathbf{x}_1), ..., P(\mathbf{x}_n)\}$. We define $g(\mathbf{x})^T := (1, \tilde{v}, \gamma)$, while $\mathbf{G}$ is the $n \times 3$ regression matrix whose ith row corresponds to $g(\mathbf{x}_i)^T$. The $n \times n$ diagonal matrix $\mathbf{K}$ contains the kernel weights obtained via the AMK kernel structure. Including $\tilde{v}$ as input to $\mathbf{G}$ is motivated by its unique importance in estimating power output, while the inclusion of γ allows us to capture its localized impact on power output, conditional on the environmental conditions within its neighborhood.

Algorithm 1 The YAMK method for Power Curve Modeling

Input training data $\mathcal{D}^{tr} = \{\tilde{v}_i, \theta_i, \rho_i, I_i, \gamma_i, P_i\}_{i=1}^n$
Input target inputs $\mathcal{D}^{ts} = \{\tilde{v}_s^*, \theta_s^*, \rho_s^*, I_s^*, \gamma_s^*\}_{s=1}^S$
Determine the kernel parameters λ_v, ν_θ, λ_ρ, and λ_I.
  for $s \in \{1, ..., S\}$ **do**
    Construct $\mathbf{K}_1$ and $\mathbf{K_2}$ using the AMK kernel weights, such that their ith diagonal entries are computed as $K(\tilde{v}_s^*, \tilde{v}_i)K(\theta_s^*, \theta_i)K(\rho_s^*, \rho_i)$ and $K(\tilde{v}_s^*, \tilde{v}_i)K(\theta_s^*, \theta_i)K(\rho^*, I_i)$, respectively.
    Set $g(\mathbf{x}_s^*) = \{1, \tilde{v}_s^*, \gamma_s^*\}$, and $\mathbf{G}$ as the correspondent $n \times 3$ training input matrix, for which the ith row is defined as $\{1, \tilde{v}_i, \gamma_i\}$.
    Estimate $\hat{f}_1(\mathbf{x}_s^*)$ and $\hat{f}_2(\mathbf{x}_s^*)$ using (7).
    Predict at $\mathbf{x}_s^* = \{\tilde{v}_s^*, \theta_s^*, \rho_s^*, I_s^*, \gamma_s^*\}$ using (9).
  end for
return the final yaw-adjusted power curve, defined by $\hat{\mathbf{P}} = \{\hat{f}(\tilde{v}_s^*, \theta_s^*, \rho_s^*, I_s^*, \gamma_s^*)\}_{s=1}^S$

Algorithm 1 summarizes the framework of the YAMK method. At our disposal is a training dataset, denoted as $\mathcal{D}^{tr} = \{\mathbf{x}_i, \gamma_i, P(\mathbf{x}_i)\}_{i=1}^n$, and a set of target inputs, denoted by $\mathcal{D}^* = \{\mathbf{x}_s^*, \gamma_s^*\}_{s=1}^S$, for which a set of predictions are needed, $\{\hat{f}(\mathbf{x}_s^*, \gamma_s^*)\}_{s=1}^S$. For each variable in $\mathbf{x}$, we first find the set of univariate kernel bandwidth parameters. For wind speed, air density, and turbulence intensity, we use univariate Gaussian kernels. Since wind direction

is a circular variable, we use the Von Mises kernel, which is defined as in (8).

$$K_\nu(\theta^*, \theta_i) = \frac{\exp\{\nu \cos(\theta^* - \theta_i)\}}{2\pi I_0(\nu)}, \tag{8}$$

where $I_0(\cdot)$ is the modified Bessel function of order 0, and ν is the concentration parameter. The parameters for all the kernels, namely λ_v, ν_θ, λ_ρ, and λ_I are found using the direct plug-in (DPI) method [34, 35].

Once the kernel parameters have been estimated, we estimate the coefficients of the locally weighted polynomial regression models using (7), which are then plugged in to make a prediction $\hat{f}(\mathbf{x}^*)$ as in (9).

$$\hat{f}(\mathbf{x}^*) = \frac{1}{2}\left[\hat{f}_1(\tilde{v}^*, \theta^*, \rho^*) + \hat{f}_2(\tilde{v}^*, \theta^*, I^*)\right]. \tag{9}$$

4. Results and Discussions

We test the performance of the YAMK method on the dataset described in Section 2. Two case studies are presented, one for each wind turbine (Turbine ID#: R80711 and R80721). We begin by describing the benchmarks, the evaluation procedure, then present the numerical results and analyses.

4.1. Benchmarks

We compare the predictive performance of the YAMK approach to a number of power curve modeling benchmarks, listed below as B1-B5:

(B1) The Method of Bins (BIN$_{v,\rho}$) is a non-parametric approach for power curve modeling described in the IEC 61400-12-1 standard [36]. It discretizes the density-corrected wind speed domain into a number of bins, each with bin width of 0.5 m/s, and then takes the average of the power values within each bin as the estimated power, as expressed in (10).

$$\bar{v}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} \tilde{v}_{ik} \quad \text{and} \quad \bar{P}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} P(\bar{v}_{ik}), \tag{10}$$

where n_k is the number of data points within the kth wind speed bin, while $\bar{v}_k$ and $\bar{P}_k$ are the average wind speed and power output within the kth bin, respectively. The prediction for a target input, $\mathbf{x}_s^*$, is obtained as in (11).

$$\hat{f}(\mathbf{x}_s^*) = \sum_{k=1}^{\mathcal{K}} \mathbb{I}(\tilde{v}_s^* \in \mathcal{B}_k)\bar{P}_k, \tag{11}$$

where $\mathcal{K}$ is the total number of bins, $\mathbb{I}(\cdot)$ is the indicator function, and $\mathcal{B}_k$ is the set of density-corrected wind speed values within the kth bin.

(B2-B3) k nearest neighbors (kNN): kNN is widely used for power curve modeling [7]. It computes a local prediction at the s-th target input as the average of the outputs of its k-nearest neighbors, as expressed in (12).

$$\hat{f}(\mathbf{x}_s^*) = \frac{1}{k} \sum_{\mathbf{x}_i \in \mathcal{N}(\mathbf{x}_s^*)} P(\mathbf{x}_i). \tag{12}$$

We implement two variants of kNN. The first variant constructs a one-dimensional power curve by solely using the density-corrected wind speed as input, that is, in (12), we set $\mathbf{x} := \{\tilde{v}\}$. We call this method as $\text{kNN}_{v,\rho}$ because it takes wind speed and air density (through the density correction) as inputs. The second variant, which we call $\text{kNN}_{v,\rho,\gamma}$ directly integrates yaw misalignment γ as an additional input beside the density-corrected wind speed, that is, in (12), we set $\mathbf{x} := \{\tilde{v}, \gamma\}$.

We implement kNN using the `caret` package in the statistical programming language `R`. The number of nearest neighbors, k, has a significant impact on the final predictive performance, and is selected by heuristically searching for the value of k that minimizes the out-of-sample error [37].

(B4-B5) Additive Multivariate Kernel (AMK): This is the kernel-based regression method proposed by Lee *et al.* [23] and explained in Section 3.1. Two variants of AMK are implemented. The first variant, denoted by $\text{AMK}_{v,\theta,\rho,I}$, is the same model proposed in [23], and uses $\mathbf{x} := \{\tilde{v}, \theta, \rho, I\}$ in (6). Its kernel structure comprises the addition of two trivariate kernels (one for $\{\tilde{v}, \theta, \rho\}$, and another for $\{\tilde{v}, \theta, I\}$), that are obtained as the product of three univariate kernels. This variant does not take yaw error into account, and unlike YAMK, computes the prediction as a local neighborhood average. The second variant, $\text{AMK}_{v,\theta,\rho,I,\gamma}$, augments this input set by γ, such that $\mathbf{x} := \{\tilde{v}, \theta, \rho, I, \gamma\}$ in (6). Its kernel structure comprises the addition of three trivariate kernels (one for $\{\tilde{v}, \theta, \rho\}$, one for $\{\tilde{v}, \theta, I\}$, and one for for $\{\tilde{v}, \theta,$ and $\gamma\}$), that are formed by the product of three univariate kernels.

4.2. Evaluation and Results

For evaluation, we implement K-fold cross validation, where we set the number of folds $K = 5$, as is typical in machine learning practice. Predictions are evaluated using the normalized root mean squared error (NRMSE),

Table 1: 5-fold average NRMSE of power curve models for turbines R80711 and R80721. Bold-faced values indicate best performance. Percentage improvements, denoted as % IMP, are computed relative to the best-performing method.

Method	**Number of Inputs**	**Turbine 1 (R08711)**	**Turbine 2 (R80721)**	**Avg.**	**% IMP**
$\text{BIN}_{v,\rho}$	2	2.585	2.260	2.423	32.13%
$\text{kNN}_{v,\rho}$	2	2.078	1.713	1.896	13.27%
$\text{kNN}_{v,\rho,\gamma}$	3	2.193	1.729	1.961	16.16%
$\text{AMK}_{v,\theta,\rho,I}$	4	1.827	1.570	1.699	3.21%
$\text{AMK}_{v,\theta,\rho,I,\gamma}$	5	1.890	1.588	1.739	5.46%
$\text{YAMK}_{v,\theta,\rho,I,\gamma}$	5	**1.780**	**1.508**	**1.644**	-

expressed as in (13). For S_k testing data points within the kth fold, the NRMSE of a power curve model $\mathcal{M}$, is given by:

$$\text{NRMSE}(\mathcal{M}) = \frac{1}{K}\sum_{k=1}^{K}\left[\frac{100\%}{p_r}\sqrt{\frac{\sum_{s=1}^{S_k}[P(\mathbf{x}_s^*) - \hat{f}(\mathbf{x}_s^*)]^2}{S_k}}\right], \tag{13}$$

where P_r refers to the rated power. For both R80711 and R80721, the rated power is $P_r = 2050$ kW.

Table 1 and Figure 5 show the averages and boxplots of the 5-fold NRMSE, respectively, across all methods for turbines R80711 and R80721. Looking at Table 1 and Figure 5, we can draw few key insights. First, one of the key findings is that a direct integration of yaw misalignment as an additional input to a multivariate power curve model does not lead to any improvement in the predictive performance (if any, it actually leads to a slight deterioration). This is evident by how $\text{kNN}_{v,\rho,\gamma}$ and $\text{AMK}_{v,\theta,\rho,I,\gamma}$ perform worse than $\text{kNN}_{v,\rho}$, and $\text{AMK}_{v,\theta,\rho,I}$, respectively.

The YAMK method, on the other hand, outperforms all of the benchmarks, with percentage improvements, on average, reaching up to 32.13% over the method of bins, and at least 3.21% relative to its closest competitor, $\text{AMK}_{v,\theta,\rho,I}$. The improvements appear to be more pronounced for turbine R80721 than R80711, which may be partly explained by the higher-on-average yaw errors for R80721 relative to R80711. Hence, the merit of leveraging the yaw-to-power correlations becomes more substantial.

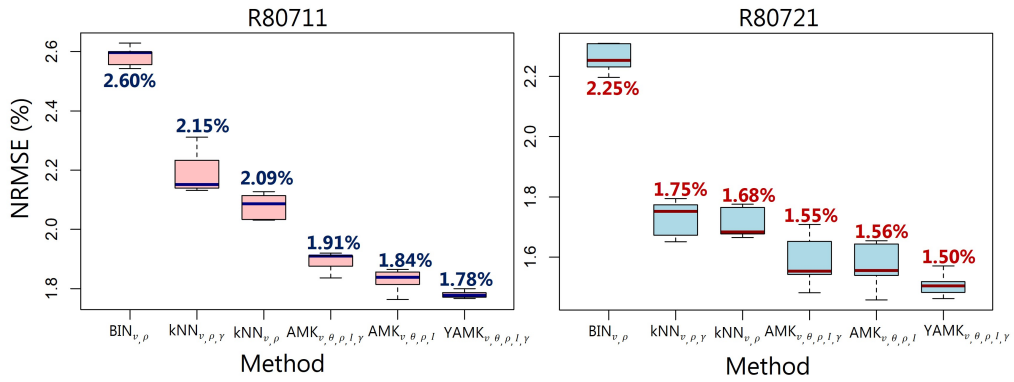

Figure 5: Boxplots of the NRMSE values across the five folds, for different methods. The numbers on top of the boxplots denote the median NRMSE value.

Figure 6 shows the estimated power curves for R80711 (left) and R80721 (right) separated into two groups: Group 1 (red circles) represents power predictions for testing observations that had yaw errors below the median yaw misalignment value (5.17° for R80711 and 5.51° for R80721), while Group 2 represents those for which the yaw errors were higher than the median values. In both R80711 and R80721, the predicted power curves for the first group's data appear to dominate those for the second group, suggesting that YAMK effectively takes into account the influence of yaw errors on power predictions. The power gains (in %) of Group 1 relative to Group 2 are 1.73% and 3.18% for Turbines R80711 and R80721, respectively, which aligns with what has been shown in the literature [30].

The distribution of the prediction errors of the wind power reveals the accuracy of the prediction approach, wherein a power curve model with poor accuracy will have its prediction errors distributed widely, as opposed to a sharper, concentrated distribution for an accurate power curve model. Figure 7 shows the probability density distributions of the wind power prediction errors for three models: $\text{BIN}_{v,\rho}$, $\text{AMK}_{v,\theta,\rho,I.\gamma}$, and $\text{YAMK}_{v,\theta,\rho,I,\gamma}$. As we can observe, the distribution of the prediction errors from $\text{BIN}_{v,\rho}$ is relatively spread out. AMK's distribution of prediction errors is better than that of $\text{BIN}_{v,\rho}$ highlighting the merit of integrating additional environmental variables. YAMK's error distribution is the sharpest among the three models, suggesting the importance of effectively integrating yaw as an additional operational input. This is further confirmed in Figure 7(b) and (d) which show

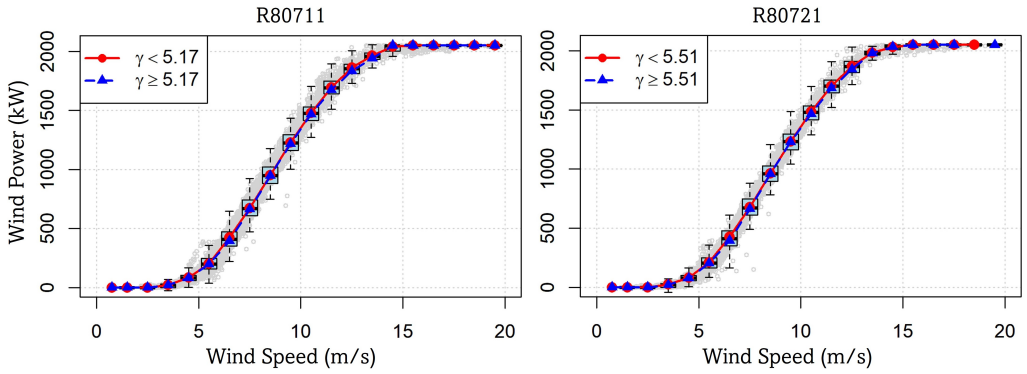

Figure 6: Final power curves for R80711 (left) and R80721 (right), separated into two groups: Group 1 (red circles): yaw errors below the median yaw misalignment value (5.17° for R80711 and 5.51° for R80721); Group 2 (blue squares): yaw errors above the median yaw misalignment value (5.17° for R80711 and 5.51° for R80721). Predicted power gains (in %) between the two groups are 1.73% and 3.18% for R80711 and R80721, respectively.

the actual versus predicted wind power values from $\text{YAMK}_{v,\theta,\rho,I,\gamma}$, which closely match with the 45° line, indicating closeness to a perfect prediction.

Finally, we would like to assess the performance of YAMK across different wind speed values. Figure 8 shows the median absolute error (scaled by P_r) versus the corresponding wind speed values (wind speed bins of 1 m/s are used), for four methods: $\text{YAMK}_{v,\theta,\rho,I,\gamma}$ (black circles), $\text{AMK}_{v,\theta,\rho,I,\gamma}$ (red triangles), $\text{kNN}_{v,\rho}$ (green squares), and $\text{BIN}_{v,\rho}$ (blue diamonds). Errors from both turbines have been pooled to produce this analysis. Again, we can immediately see that YAMK significantly improves over all methods across almost all wind speed ranges. A closer look reveals that YAMK's largest improvements over its closest competitor, AMK, manifests itself in the early and middle portions of Region II of a turbine's power curve (between 3.5 and 11 m/s). As the wind speed approaches the rated power region (Region III), the performance of YAMK gradually converges towards that of AMK (especially noticeable at wind speeds higher than 12 m/s). This finding is in line with the analysis in Figure 4 which suggests that the impact of yaw on power is more articulated for the early and middle portions of Region II in a turbine's power curve. We also note that the relative improvements of YAMK over methods that do not take the yaw as an input in first place (namely, $\text{kNN}_{v,\rho}$ and $\text{BIN}_{v,\rho}$) are maintained across all wind speed values.

The analysis above shows how our yaw-adjusted power curve is able to

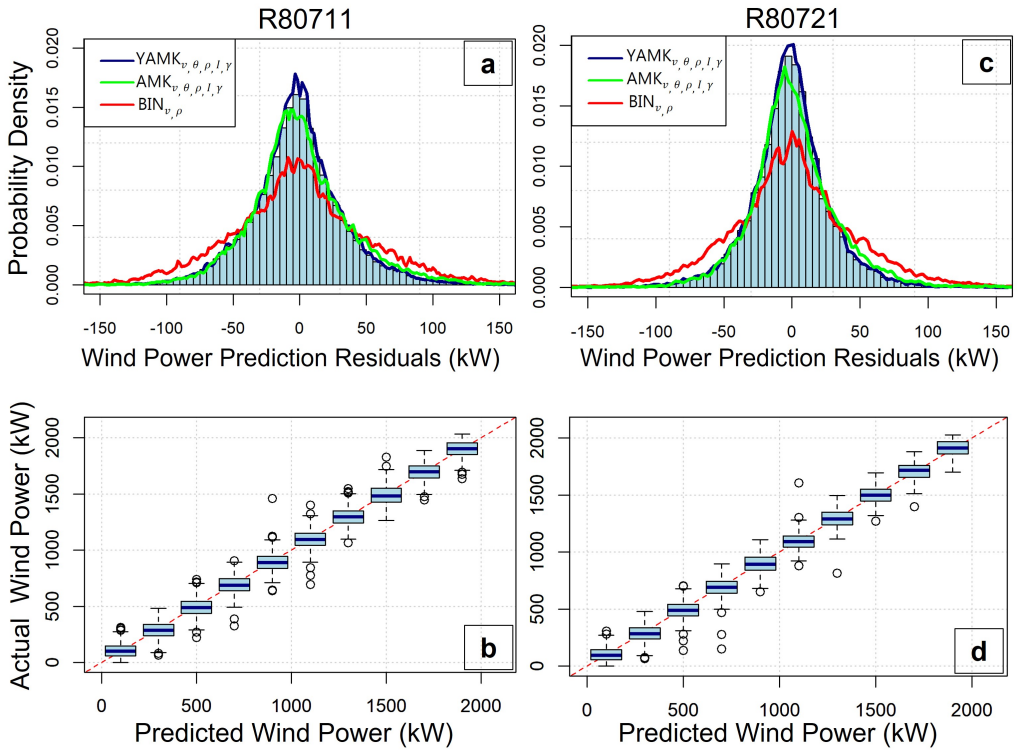

Figure 7: (a) and (c): Residuals of wind power prediction for three models: $\text{BIN}_{v,\rho}$ (red), $\text{AMK}_{v,\theta,\rho,I,\gamma}$ (green), and $\text{YAMK}_{v,\theta,\rho,I,\gamma}$ (blue), on top of the histograms of YAMK's residuals. (b) and (d): Actual versus predicted wind power values (in kW) for the YAMK method, lying closely to the 45° line suggesting high-quality predictive performance.

(1) leverage the AMK special kernel structure to scalably integrate several environmental parameters, and (2) effectively harness the yaw-to-power correlations in order to make significant predictive improvements, on top of those provided by the AMK and other nonparametric power curve models. Those results affirm our hypothesis: yaw is indeed a significant predictor of power, but its significance may only be statistically leveraged in a power curve model once *conditioned on* a neighborhood of environmental parameters.

5. Conclusions

In this paper, we proposed the notion of a "yaw-adjusted power curve." Instead of directly adding yaw error as an additional input to a multivariate

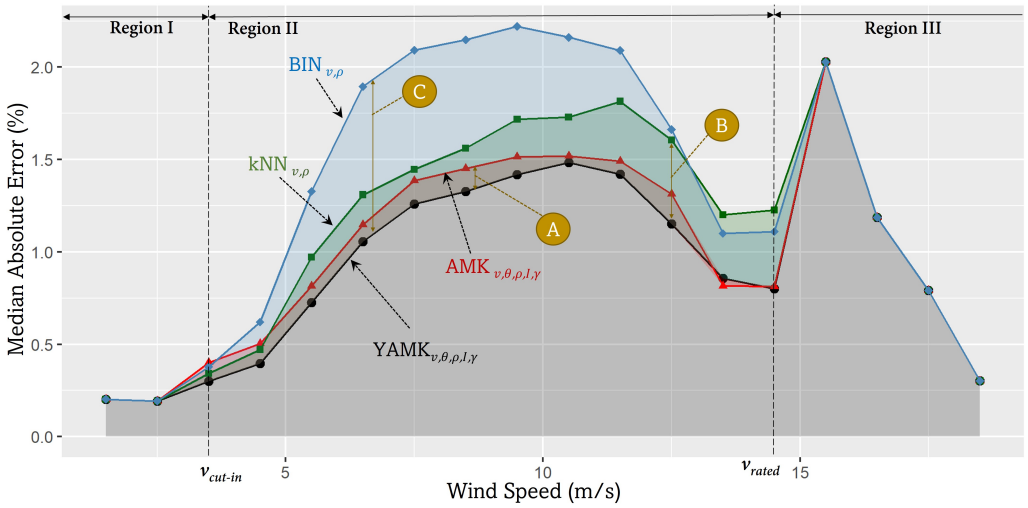

Figure 8: Performance of YAMK$_{v,\theta,\rho,I,\gamma}$ (black circles), AMK$_{v,\theta,\rho,I,\gamma}$ (red triangles), kNN$_{v,\rho}$ (green squares), and BIN$_{v,\rho}$ (blue diamonds), in terms of median absolute error (scaled by P_r). Points (A), (B), and (C) show the maximum differences in performance of each method relative to YAMK$_{v,\theta,\rho,I,\gamma}$. Those correspond to percentage improvements of 9.35% (point A), 28.3% (point B), and 44.3% (point C), respectively.

power curve model, which is found not to result in any competitive advantage, our approach learns the local yaw-to-power relationship conditional on an effective neighborhood of environmental variables defined using a special kernel structure which was recently proposed in the literature. Our experiments on a real-world dataset suggest that such approach results in significant improvements, in terms of power output estimation accuracy, relative to several existing power curve models.

We believe that the proposed yaw-adjusted power curve model can be valuable to several wind farm operations that primarily rely on wind power curves such as wind power prediction, asset monitoring and prognostics, maintenance scheduling, among others. In addition, such yaw-adjusted wind power curves can be highly relevant to the emerging area of yaw-based power production optimization in wind farms. Recent studies suggest that yaw-based production optimization can significantly improve power production. A pivotal assumption in such optimization routines, however, is the perfect knowledge about how yaw errors impact the power output of a wind turbine. Without a reliable method to accurately estimate the yaw-to-power response surface, as the one proposed herein, the power gains from such emerging

and promising efforts may be forfeited or largely compromised. An area of future research is to explore the impact of the improvement in power curve estimation from our model on such power production optimization routines.

Another interesting line of future research is to consider turbine-to-turbine dependencies. It is known that a turbine's yaw error does not only impact its power output, but also that of its neighbors in a wind farm. In this research, we have only considered a period of time when the turbines under study were upstream relative to the incoming winds. Future research may therefore look into modeling turbine dependencies when constructing the yaw-to-power relationship.

Data Availability

The wind turbine dataset used in this work has been made publicly available by Engie through its website and can be accessed at `https://opendata-renewables.engie.com/` [27]. Air pressure measurements, co-located with Engie's dataset, have been obtained from Meteoblue [29].

Acknowledgment

This research has been partly supported by the Rutgers Research Council Grant Program, and by the National Science Foundation (NSF) under Grant ECCS-2114422.

References

[1] C. Sweeney, R. J. Bessa, J. Browell, P. Pinson, The future of forecasting for renewable energy, Wiley Interdisciplinary Reviews: Energy and Environment 9 (2) (2020) e365.

[2] A. Ezzat, Turbine-specific short-term wind speed forecasting considering within-farm wind field dependencies and fluctuations, Applied Energy 269 (2020) 115034.

[3] C. Carrillo, A. O. Montaño, J. Cidrás, E. Díaz-Dorado, Review of power curve modelling for wind turbines, Renewable and Sustainable Energy Reviews 21 (2013) 572–581.

[4] M. Lydia, S. S. Kumar, A. I. Selvakumar, G. E. P. Kumar, A comprehensive review on wind turbine power curve modeling techniques, Renewable and Sustainable Energy Reviews 30 (2014) 452–460.

[5] Y. Wang, Q. Hu, L. Li, A. M. Foley, D. Srinivasan, Approaches to wind power curve modeling: A review and discussion, Renewable and Sustainable Energy Reviews 116 (2019) 109422.

[6] B. Niu, H. Hwangbo, L. Zeng, Y. Ding, Evaluation of alternative power production efficiency metrics for offshore wind turbines and farms, Renewable Energy 128 (2018) 81–90.

[7] A. Kusiak, H. Zheng, Z. Song, On-line monitoring of power curves, Renewable Energy 34 (6) (2009) 1487–1493.

[8] P. Pinson, Wind energy: Forecasting challenges for its operational management, Statistical Science 28 (4) (2013) 564–585.

[9] A. Ezzat, M. Jun, Y. Ding, Spatio-temporal asymmetry of local wind fields and its impact on short-term wind forecasting, IEEE Transactions on Sustainable Energy 9 (3) (2018) 1437–1447.

[10] Y. He, A. Kusiak, Performance assessment of wind turbines: data-derived quantitative metrics, IEEE Transactions on Sustainable Energy 9 (1) (2017) 65–73.

[11] P. Guo, D. Infield, Wind turbine power curve modeling and monitoring with Gaussian process and SPRT, IEEE Transactions on Sustainable Energy 11 (1) (2018) 107–115.

[12] P. Papadopoulos, D. W. Coit, A. A. Ezzat, STOCHOS: Stochastic opportunistic maintenance scheduling for offshore wind farms, arXiv preprint arXiv:2207.02274 (2022).

[13] P. Papadopoulos, D. Coit, A. Ezzat, Seizing opportunity: Maintenance optimization in offshore wind farms considering accessibility, production, and crew dispatch, IEEE Transactions on Sustainable Energy (2021).

[14] P. A. Fleming, A. Scholbrock, A. Jehu, S. Davoust, E. Osler, A. D. Wright, A. Clifton, Field-test results using a nacelle-mounted lidar for improving wind turbine power capture by reducing yaw misalignment,

in: Journal of Physics: Conference Series, Vol. 524, IOP Publishing, 2014, p. 012002.

[15] V. Sohoni, S. Gupta, R. Nema, A critical review on wind turbine power curve modelling techniques and their applications in wind based energy systems, Journal of Energy 2016 (2016).

[16] M. G. Khalfallah, A. M. Koliub, Suggestions for improving wind turbines power curves, Desalination 209 (1-3) (2007) 221–229.

[17] B. Stephen, S. J. Galloway, D. McMillan, D. C. Hill, D. G. Infield, A copula model of wind turbine performance, IEEE Transactions on Power Systems 26 (2) (2010) 965–966.

[18] V. Thapar, G. Agnihotri, V. K. Sethi, Critical analysis of methods for mathematical modelling of wind turbines, Renewable Energy 36 (11) (2011) 3166–3177.

[19] A. Marvuglia, A. Messineo, Monitoring of wind farms' power curves using machine learning techniques, Applied Energy 98 (2012) 574–583.

[20] S. Shokrzadeh, M. J. Jozani, E. Bibeau, Wind turbine power curve modeling using advanced parametric and nonparametric methods, IEEE Transactions on Sustainable Energy 5 (4) (2014) 1262–1269.

[21] T. Ouyang, A. Kusiak, Y. He, Modeling wind-turbine power curve: A data partitioning and mining approach, Renewable Energy 102 (2017) 1–8.

[22] R. K. Pandit, D. Infield, J. Carroll, Incorporating air density into a Gaussian process wind turbine power curve model for improving fitting accuracy, Wind Energy 22 (2) (2019) 302–315.

[23] G. Lee, Y. Ding, M. G. Genton, L. Xie, Power curve estimation with multivariate environmental factors for inland and offshore wind farms, Journal of the American Statistical Association 110 (509) (2015) 56–67.

[24] B. Golparvar, P. Papadopoulos, A. Ezzat, R.-Q. Wang, A surrogate-model-based approach for estimating the first and second-order moments of offshore wind power, Applied Energy 299 (2021) 117286.

[25] R. K. Pandit, D. Infield, A. Kolios, Gaussian process power curve models incorporating wind turbine operational variables, Energy Reports 6 (2020) 1658–1669.

[26] D. Astolfi, Wind turbine operation curves modelling techniques, Electronics 10 (3) (2021) 269.

[27] Engie's Open Data, https://opendata-renewables.engie.com/.

[28] T. Duc, G. Giebel, T. Göçmen, M. Korpås, O. Coupiac, Optimization of wind farm power production using innovative control strategies (2017).

[29] Meteoblue Official Website, https://www.meteoblue.com/en/weather/week/passaic_united-states-of-america_5102443.

[30] K. A. Kragh, M. H. Hansen, Potential of power gain with improved yaw alignment, Wind Energy 18 (6) (2015) 979–989.

[31] J. Jeon, J. W. Taylor, Using conditional kernel density estimation for wind power density forecasting, Journal of the American Statistical Association 107 (497) (2012) 66–79.

[32] E. A. Nadaraya, On estimating regression, Theory of Probability & Its Applications 9 (1) (1964) 141–142.

[33] Y. Ding, Data Science for Wind Energy, CRC Press, 2019.

[34] D. Ruppert, S. J. Sheather, M. P. Wand, An effective bandwidth selector for local least squares regression, Journal of the American Statistical Association 90 (432) (1995) 1257–1270.

[35] C. C. Taylor, Automatic bandwidth selection for circular density estimation, Computational Statistics & Data Analysis 52 (7) (2008) 3493–3500.

[36] Wind Energy Generation Systems - Part 12-1: Power Performance Measurements of Electricity Producing Wind Turbines, IEC 61400-12-1International Electrotechnical Commission (2017).

[37] M. Kuhn, Building predictive models in R using the caret package, Journal of Statistical Software 28 (2008) 1–26.