Power Curve Estimation with Multivariate Environmental Factors for Inland and Offshore Wind Farms

Giwhyun Lee

Korea Army Academy

Yeongcheon, Republic of Korea

email: giwhyunlee@gmail.com

Yu Ding

Industrial and Systems Engineering,

Texas A&M University, TX, USA

email: yuding@iemail.tamu.edu

Marc G. Genton

CEMSE Division,

King Abdullah University of Science and Technology, Saudi Arabia

email: marc.genton@kaust.edu.sa

Le Xie

Electrical and Computer Engineering,

Texas A&M University, TX, USA

email: Lxie@ece.tamu.edu

Abstract

In the wind industry, a power curve refers to the functional relationship between the power output generated by a wind turbine and the wind speed at the time of power generation. Power curves are used in practice for a number of important tasks including predicting wind power production and assessing a turbine's energy production efficiency. Nevertheless, actual wind power data indicate that the power output is affected by more than just wind speed. Several other environmental factors, such as wind direction, air density, humidity, turbulence intensity, and wind shears, have potential impact. Yet, in industry practice, as well as in the literature, current power curve models primarily consider wind speed and, sometimes, wind speed and direction. We propose an additive multivariate kernel method that can include the aforementioned environmental factors as a new power curve model. Our model provides, conditional on a given environmental condition, both the point estimation and density estimation of power output. It is able to capture the nonlinear relationships between environmental factors and the wind power output, as well as the high-order interaction effects among some of the environmental factors. Using operational data associated with four turbines in an inland wind farm and two turbines in an offshore wind farm, we demonstrate the improvement achieved by our kernel method.

KEYWORDS: Additive multivariate kernel regression, Nonparametric estimation, Turbine performance assessment, Wind power forecast

1. INTRODUCTION

Wind energy is one of the fastest growing renewable energy sources. According to a report issued by the American Wind Energy Association (AWEA), wind power installation in the U.S. increased by more than a factor of ten in the past decade, from 4,232 megawatts (MW) in 2001 to 46,919 MW by the end of 2011 (AWEA 2012). The U.S. Department of Energy advocates working towards the goal that wind power accounts for 20 percent of the total electricity generated in the U.S. by 2030 (DOE 2008). To manage wind turbines and to plan wind energy production, it is critical to assess wind power generation under a given weather profile. The so-called power curve plays a central role in this task (Giebel, Brownsword, Kariniotakis, Denhard and Draxl 2011; Monteiro, Bessa, Miranda, Botterud, Wang and Conzelmann 2009). In the wind industry, the power curve measures the relationship between power output of a turbine on the wind speed. In this paper, we estimate the power curve associated with individual turbines at both inland and offshore wind farms using turbine-specific power output data and environmental data measured from a meteorological mast on the corresponding farm.

We first explain the basics of the power curve. Denote by y the power output from a wind turbine and by $\boldsymbol{x}$ the vector of explanatory variables. V is the wind speed. In wind power production, as illustrated in the left panel of Figure 1, a turbine starts to produce power after the wind reaches the cut-in speed, V_{ci}. A nonlinear relation between y and V then ensues, until wind reaches the rated wind speed, V_r. When the wind speed is beyond V_r, the turbine's output power will be restricted at the rated power output, y_r, also known as the nominal power capacity of the turbine, using control mechanisms such as pitch control and rotor speed regulation. The turbine will be halted when the wind reaches the cut-out speed, V_{co}, because high wind is deemed harmful to the safety of a turbine. For the power curve shown in the left panel of Figure 1, $\boldsymbol{x} := (V)$.

The wind industry makes use of power curves for at least two important purposes. The

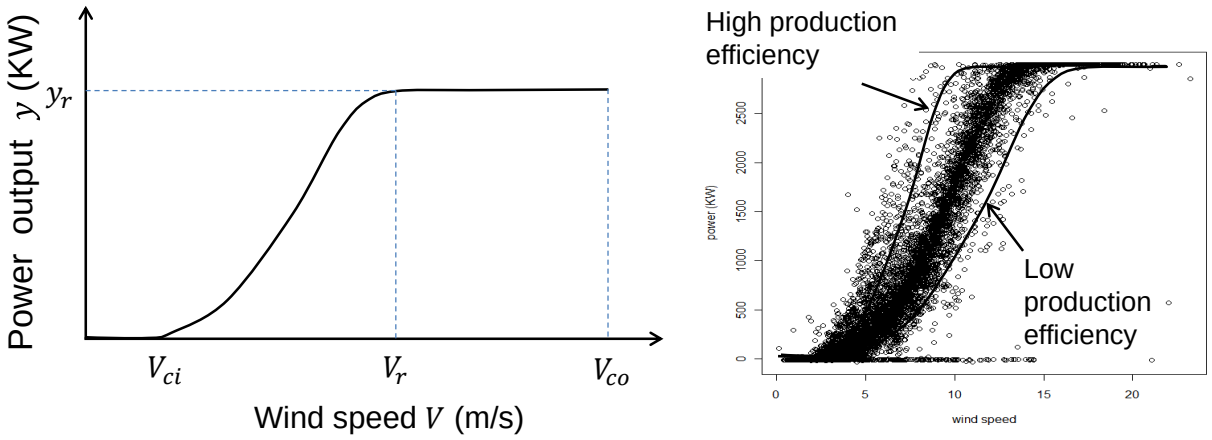


Figure 1: An example of a power curve: V_{ci} is the cut-in wind speed, V_{co} is the cut-out wind speed, V_r is the rated wind speed, and y_r is the corresponding rated power output. In the right panel where real power production data are shown, the power outputs are normalized by the rated power output, to protect the identity of the turbine manufacturer. The same treatment is applied to all power curve plots throughout the paper.

first is to forecast wind power (Giebel et al. 2011; Monteiro et al. 2009) in two steps. First, wind speeds are forecast and then this forecast is converted to a power forecast using a power curve. The second purpose of power curves is for turbine performance assessment and turbine health monitoring (Albers, Klug and Westermann 1999; Uluyol, Parthasarathy, Foslien and Kim 2011; Stephen, Galloway, McMillan, Hill and Infield 2011), in which a power curve is used to characterize a turbine's power production efficiency by noting the changes in the position and slope of the turbine's power curve; for an illustration, see Figure 1, right panel. We note that wind forecasting is beyond the focus of this paper; it is a subfield of its own; for details on wind forecasting, interested readers should refer to Gneiting, Larson, Westrick, Genton and Aldrich (2006), Monteiro et al. (2009), Hering and Genton (2010), and Zhu and Genton (2012).

The curve shown in the left panel of Figure 1 is an ideal power curve, also known as the nominal power curve, typically provided by a turbine manufacturer. The right panel of Figure 1 shows the actual power output and wind measurement data associated with a

turbine, which presents a more complicated picture. Even though the general trend shown in the data tends to agree with the nominal power curve, there appears to be a considerable amount of information that cannot be accounted for by a simple V-versus-y curve. Between 5 meters per second (m/s) and 15 m/s, there are large amounts of power data for any given wind speed. What this implies is that if V is used as the sole explanatory variable, the prediction of wind power suffers a high degree of uncertainty. We investigate whether, and how, more explanatory variables can be included to make a better fit to the power data.

In fact, the meteorological mast on each wind farm included a wide array of sensors that measure more than just wind speed. Other environmental variables measured include wind direction, D, temperature, T, air pressure, P, and humidity, H. Based on the wind speed measurements, it is also possible to calculate turbulence intensity, I (equal to the standard deviation of short-duration wind speeds divided by the average wind speed of the same duration) and wind shear, S (using wind speeds measured at different heights). If we expand our input variable set to include these environmental factors, we could have $\boldsymbol{x} := (V, D, T, P, H, I, S)$. Then, our technical objective would be to estimate the conditional density, $p(y|\boldsymbol{x})$, or the conditional expectation, $E(y|\boldsymbol{x})$. Technically, $E(y|\boldsymbol{x})$ is no longer a power curve when $\boldsymbol{x}$ includes multiple elements; it becomes a power response surface. For the sake of being consistent with industrial convention, we use the term "power curve" in its broad meaning, covering the cases of both one-dimensional power curves and multi-dimensional power response surfaces.

The current industrial practice of estimating the power curve relies on a nonparametric approach, known as the binning method, recommended by the International Electrotechnical Commission (IEC 2005). The basic idea of the binning method is to discretize the domain of wind speed into a finite number of bins, say, using a bin width of 0.5 m/s. Then, the value to be used for representing the power output for a given bin is simply the sample average of

all the data points falling within that specific bin, namely:

$$y_i = \frac{1}{N_i} \sum_{j=1}^{N_i} y_{i,j}, \tag{1}$$

where $y_{i,j}$ is the power output of the j^{th} data point in bin i, and N_i is the number of data points in bin i. In the binning method, almost all other environmental variables are ignored, except for the so-called air density adjustment, for which we will present a detailed expression later.

Many existing methods of fitting a power curve are similar to the binning method in the sense that only wind speed is used as the sole explanatory variable, although the specific techniques used for curve fitting were quite different (Yan, Osadciw, Benson and White 2009; Kusiak, Zheng and Song 2009; Uluyol et al. 2011; Osadciw, Yan, Ye, Benson and White 2010; Hayes, Ilie, Porpodas, Djokic and Chicco 2011). For instance, Yan et al. (2009) and Osadciw et al. (2010) used a symmetric sigmoid function and a Gaussian cumulative density function for curve fitting, and Kusiak et al. (2009) used a logistic function. These methods are of parametric flavor. Kusiak et al. (2009) also suggested a nonparametric approach, which is to use the k-nearest neighbor (k-NN) method to make power predictions.

Wan, Ela and Orwig (2010) extended the binning method. In one aspect, they studied the wind direction effect, but their approach was simply to divide wind direction into a few disjointed sub-directions; doing this is, in fact, an action of binning. Another extension is that they tried a neural network model that took both wind speed and air density as inputs. However, their study concluded that doing so does not appear beneficial. When comparing a few different options, including curve fitting (they did not specify which curve fitting method they used) and binning, Wan et al. (2010) concluded that the binning method with air density correction produced the best power curve fitting outcome.

A handful of studies do explicitly include both wind speed and wind direction in their

models (Nielsen, Nielsen and Madsen 2002; Sanchez 2006; Pinson, Nielsen, Madsen and Nielsen 2008; Jeon and Taylor 2012). The inclusion of wind direction is not surprising because of the physical intuition that how wind blows the turbine should matter in wind power production. The specific approaches employed in these studies differed: Nielsen et al. (2002) used a local polynomial regression; Sanchez (2006) presented a dynamic combination of several prediction models based on time-varying coefficients and a recursive solution procedure; Pinson et al. (2008) used a total least squares criterion (i.e., orthogonal distance least squares), together with a Huber M-estimator, to achieve a certain degree of robustness. Kernel methods are among the sophisticated approaches that are used to model wind direction predictions (Marzio, Panzera and Taylor 2012; Marzio, Panzera and Taylor 2013; Marzio, Panzera and Taylor 2014) and model the power to wind speed/direction relationship through Jeon and Taylor (2012). Not only does Jeon and Taylor (2012) consider both wind speed and wind direction, it also produces a density estimation that can be used to account for uncertainty in wind power prediction.

It is obvious from the above literature review that despite the availability of other environmental measurements and their potential impact on power curve estimation, many current methods made use of wind speed only, while a few others used wind speed and direction. The need to develop power curve methods with multivariate dependencies has been recently noted, directly by Stephen et al. (2011) and indirectly in the studies by Tindal, Johnson, LeBlanc, Harman, Rareshide and Graves (2008) and Albers et al. (1999). Our research shows that including the extra environmental factors in a power curve model can indeed improve wind power predictions. A power curve model that incorporates multiple environmental factors also provides a useful tool for studying the relative importance of these environmental variables on wind power generation.

To fulfill the objective of developing a power curve method with multivariate dependencies, we devise an additive multivariate kernel (AMK) model. The multivariate aspect

empowers the model to capture interaction effects up to three factors, while the additive structure allows the resulting model to remain scalable as people add more explanatory environmental variables into $\boldsymbol{x}$ in the future. In the remainder of the paper, we first describe the datasets used in this study. We proceed in Section 3 to present the details of our additive multivariate kernel model. In Section 4, we compare our method with some alternative methods, arguing that the resulting kernel method produces better estimates. Finally, we end the paper with some discussion in Section 5.

2. DATA SETS

We study both in-land wind turbines (ILTs) and off-shore turbines (OSTs), and have two datasets corresponding to an inland wind farm (ILWF) and an offshore wind farm (OSWF), respectively. The datasets are denoted generally by $\mathcal{D}$ or specifically by $\mathcal{D}_{\text{ILWF}}$ or $\mathcal{D}_{\text{OSWF}}$, respectively. Table 1 summarizes the specifications of the datasets; for certain entries an approximation rather than the accurate value is given for the protection of the identities of the turbine manufacturers and wind farms.

Table 1: Specifications of the two wind farms

Wind farm	$\mathcal{D}_{\text{ILWF}}$	$\mathcal{D}_{\text{OSWF}}$
Number of meteorological masts	Multiple	Single
Number of wind turbines	200+	30+
Hub height (m)	80	70
Rotor diameter (m)	about 80	about 90
Cut-in wind speed (m/s)	3.5	3.5
Cut-out wind speed (m/s)	20	25
Rated wind speed (m/s)	around 13	around 15
Rated power (MW)	$1.5 - 2.0$	around 3
Location	In-land, U.S.	Off-shore, Europe

We choose four wind turbines and two meteorological masts from $\mathcal{D}_{\text{ILWF}}$, and two wind turbines and the single meteorological mast from $\mathcal{D}_{\text{OSWF}}$. The six turbines are denoted as WT1 to WT6, respectively, where the first four are inland turbines and the last two are

offshore ones. The environmental data in $\boldsymbol{x}$ were collected by sensors on a meteorological mast, while the power output, y, was measured at a wind turbine. Each meteorological mast has two wind turbines associated with it, meaning that the $\boldsymbol{x}$'s measured at this mast are paired with the y's of those associated turbines. For the turbines/masts layout and turbine-to-mast distances, please refer to Figure 2.

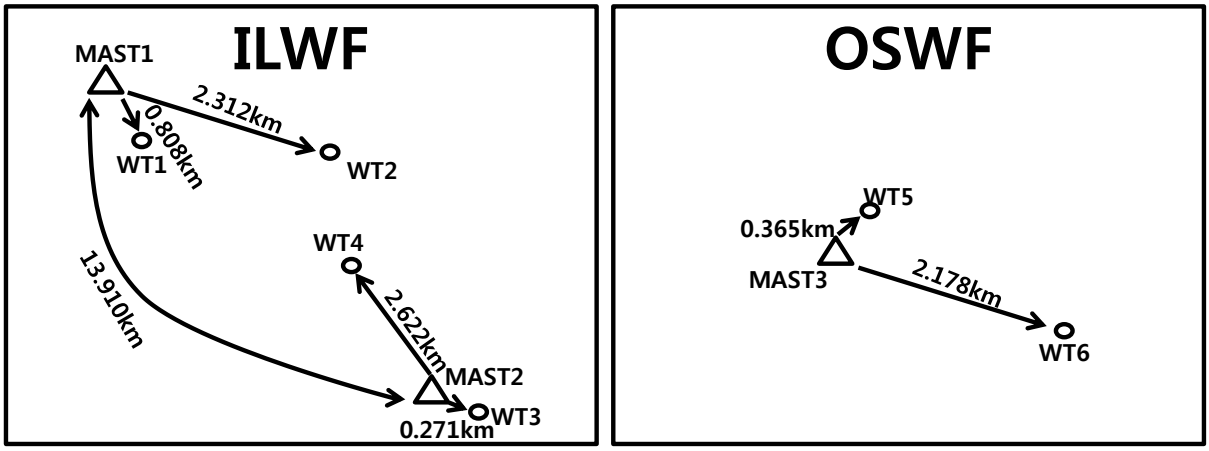


Figure 2: Layout of the turbines and masts and turbine-to-mast distances.

For WT1 and WT2 of the ILWF, the data were collected from July 30, 2010 through July 31, 2011; for WT3 and WT4 (still of the ILWF), the data were collected from April 29, 2010 through April 30, 2011, and for WT5 and WT6 of the OSWF, the data were collected from January 1, 2009 through December 31, 2009.

In current practice, data collected at wind farms are arranged in 10-minute blocks because wind speeds are considered stationary, and other environmental factors nearly constant, over a 10-minute duration. As a result, the power output, y, as well as the environmental factors, V, D, T, P, H, are the averages of the recordings in a 10-minute duration. Moreover, a few other variables can be computed as follows:

- Turbulence intensity, I: we first compute the standard deviation of the wind speeds in a 10-minute duration and denote it as $\hat{\sigma}$. Then, $I = \frac{\hat{\sigma}}{V}$, where V is the average wind

speed of the same 10-minute duration. Apparently, this turbulence intensity is similar to the coefficient of variation concept in statistics.

- Wind shear, S: wind speeds V_1 and V_2, measured at heights g_1 and g_2, respectively, are given. Then, $S = \frac{\ln(V_2/V_1)}{\ln(g_2/g_1)}$ (Rehman and Al-Abbadi 2005). For $\mathcal{D}_{\text{ILWF}}$, wind speeds are measured at two heights of 80 m and 50 m, where 80 m is the hub height. Given this instrumentation capability, one wind shear value is calculated, which is a below-hub wind shear. For $\mathcal{D}_{\text{OSWF}}$, wind speeds are measured at the heights of 116 m, 70 m, and 21 m, where 70 m is the hub height. Two wind shear values can therefore be calculated: using the 116 m/70 m pair produces an above-hub wind shear, while using the 70 m/21 m pair produces a below-hub wind shear. We denote the above-hub wind shear as S_a and the below-hub wind shear as S_b.

- Air density, ρ (kg/m^3): given air temperature, T, expressed in Kelvin and air pressure, P, expressed in Newtons/m^2, $\rho = \frac{P}{R \cdot T}$, where $R = 287$ (Joule)(kg)$^{-1}$(Kelvin)$^{-1}$ is the gas constant (Uluyol et al. 2011). In the subsequent analysis, the air density, ρ, instead of T and P, is included as an explanatory variable in $\boldsymbol{x}$. The reason is presented in the next section.

Considering the descriptions presented above, one can see that for $\mathcal{D}_{\text{OSWF}}$, there are seven explanatory variables, i.e., $\boldsymbol{x} = (V, D, \rho, H, I, S_a, S_b)$. In $\mathcal{D}_{\text{ILWF}}$, humidity measurements are not available, and the dataset has only the below-hub wind shear. Consequently, $\mathcal{D}_{\text{ILWF}}$ has five explanatory variables, namely $\boldsymbol{x} = (V, D, \rho, I, S_b)$, two fewer than $\mathcal{D}_{\text{OSWF}}$ has. Throughout this paper, by "a data point," we refer to a pair of $(\boldsymbol{x}, y)$, and we denote the total number of data points associated with a turbine as N.

Although the number of covariates in $\boldsymbol{x}$ is 5 to 7 in this study, that number can be greater with the advancement of sensor technology; for instance, if horizontal wind shear can be measured, it would add at least two more variables (left and right wind shears)

immediately to $\boldsymbol{x}$.

3. AN ADDITIVE MULTIVARIATE KERNEL METHOD FOR POWER CURVE ESTIMATION

In this section, we first provide some background information on the physical understanding of wind power generation. This physical understanding helps motivate our modeling approach undertaken subsequently.

3.1 The physics behind wind power generation

The physical law of wind power generation (Ackermann 2005; Belghazi and Cherkaoui 2012) states that:

$$y = \frac{1}{2} \cdot C_p(\beta, \lambda) \cdot \rho \cdot \pi R^2 \cdot V^3, \tag{2}$$

where R is the radius of the rotor and C_p is the so-called power coefficient, which is believed to be a function of (at least) the blade pitch angle, β, and the turbine's tip speed ratio, λ. What else might affect C_p is still a matter under investigation. Currently no formula exists to express C_p analytically in terms of its influencing factors. C_p is therefore empirically estimated and turbine manufacturers usually provide for a specific turbine its nominal power curve with the corresponding C_p values under different combinations of wind speed, V, and air density, ρ. The above expression also provides the rationale why temperature, T, and air pressure, P, are converted into air density, ρ, rather than used individually, to explain wind power.

Even though the expression in (2) on the surface suggests that the electrical power that a wind turbine extracts from the wind is proportional to V^3, an actual power curve may exhibit a different nonlinear relationship. This happens because of the tip speed ratio, $\lambda = \frac{\omega \cdot R}{V}$, where ω is the rotor speed. Consequently, C_p is also a function of wind speed, V.

The power law in (2) governs the wind power generation before the rated wind speed, V_r. The use of the pitch control mechanism levels off, and ultimately caps, the power output when it reaches the rated power output, y_r. Recall the shape of power curve shown in Figure 1. The power curve has an inflection point somewhere nearby the rated wind speed, so that the whole curve consists of a convex segment, between V_{ci} and the inflection point and a concave segment, between the inflection point and V_{co}.

Given the physical relation expressed in (2), the wind industry recognizes the need to include air density as a factor in calculating the power output, and does so through a formula known as the air density correction. If V is the raw average wind speed measured in a 10-minute duration, the air density correction is to adjust the wind speed based on the measured average air density, ρ, in the same 10-minute duration, namely

$$V' = V\left(\frac{\rho}{\rho_0}\right)^{\frac{1}{3}}, \tag{3}$$

where ρ_0 is the sea-level dry air density (=1.225 kg/m^3) per ISO atmosphere standard. The binning method with air density correction uses this corrected wind speed, V', and the power output, y, to establish a power curve. In the subsequent analysis, as well as in Section 4 where we conduct comparisons of methods, by "binning method" we refer to this air density corrected version, unless otherwise noted.

3.2 Additive multivariate kernel-based power curve model

The underlying physics of wind power generation expressed above provides some clues concerning a preferable power curve model. The following summarizes our observations:

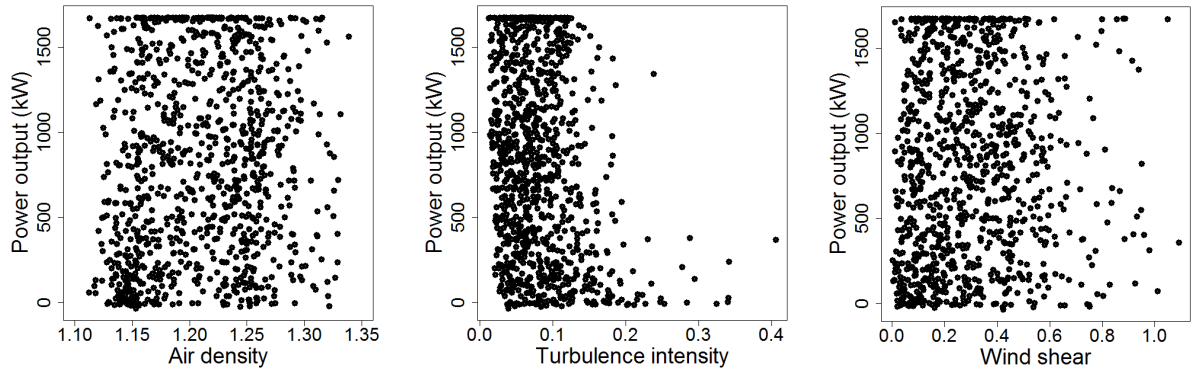
- There appear at least three important factors that affect wind power generation: wind speed, V, wind direction, D, and air density, ρ. This does not exclude the possibility that other environmental factors may also influence the power output.

- The functional relationships between the environmental factors and the power response are generally nonlinear. The complexity partially comes from the lack of understanding of C_p, which is affected by many environmental factors (V, D, and ρ included). We also stated above that there is no analytical expression linking C_p to any of the influencing factors. As a result, the functional form of a potential power curve is not known either.

- The environmental factors appear in a multiplicative relationship in the power law equation (2), indicating interactions among the factors.

To illustrate the existence of interactions among factors, we present the following plots in Figures 3 and 4. Figure 3 uses the data from $\mathcal{D}_{\text{ILWF}}$ and shows the scatter plots between the power output, y, and air density, ρ, turbulence intensity, I, and wind shear, S_b, respectively. Unconditional on wind speed, V, and wind direction, D, these environmental factors therefore have no obvious effect on the power output. Figure 4, on the other hand, presents the scatter plots between the power output and environmental factors under different wind speeds and wind directions. We do observe nonlinear relationships in these plots, and the relationships appear to be different depending on the wind conditions. This implies that interaction effects exist among wind speed, V, wind direction, D, and other environmental factors. A power curve model should characterize not only the nonlinear effects of wind speed and wind direction, but also the interaction effects among the environmental factors.

The existence of interaction effects suggests that purely additive models or generalized additive models (GAM) are unlikely to work well in modeling a power curve. There do exist sophisticated semi- or non-parametric approaches that could be capable of addressing the above identified challenges for power curve modeling, such as the Bayesian Additive Regression Trees (Chipman, George and McCulloch 2010, BART) and the Smoothing Spline ANOVA (Gu 2013, SSANOVA), as well as the kernel-based method, as used by Jeon and Taylor (2012).



(a) Air density (b) Turbulence intensity (c) Wind shear

Figure 3: Scatter plots of the power output versus environmental factors for $3.5 < V < 20, 0 < D < 360$

When we look for a modeling strategy, we settle on the kernel-based approach, namely, using a conditional kernel density estimation (CKD) (Rosenblatt 1969; Hyndman, Bashtannyk and Grunwald 1996) for estimating the condition density, $p(y|\boldsymbol{x})$, or a kernel regression (Nadaraya 1964; Watson 1964) for estimating the conditional expectation, $E(y|\boldsymbol{x})$. The reason of our choice is following. The kernel-based method appears to be a capable statistical modeling tool, not only capturing the complicated higher order interaction effects but also avoiding the need to specify a functional form of the power curve relationship. Indeed, a bivariate CKD including wind speed and direction was used by Jeon and Taylor (2012), which has produced encouraging improvement. We also believe that the aforementioned physical understanding behind wind power generation offers useful clues. We should make explicit use of them and devise a special model structure tailored to handle the power curve modeling more effectively. The simple structure of the kernel-based approach makes such tailoring easier to develop. In Section 4, we compare the results from BART and SSANOVA with that of our proposition and believe that the comparisons substantiate our claim here.

Specifically, using Rosenblatt's CKD (Rosenblatt 1969), the density of y conditional on $\boldsymbol{x}$ can be expressed as

$$\hat{f}(y|\boldsymbol{x}) = \sum_{i=1}^{N} w_i(\boldsymbol{x}) \mathcal{K}_{h_y}(y - y_i), \tag{4}$$

where

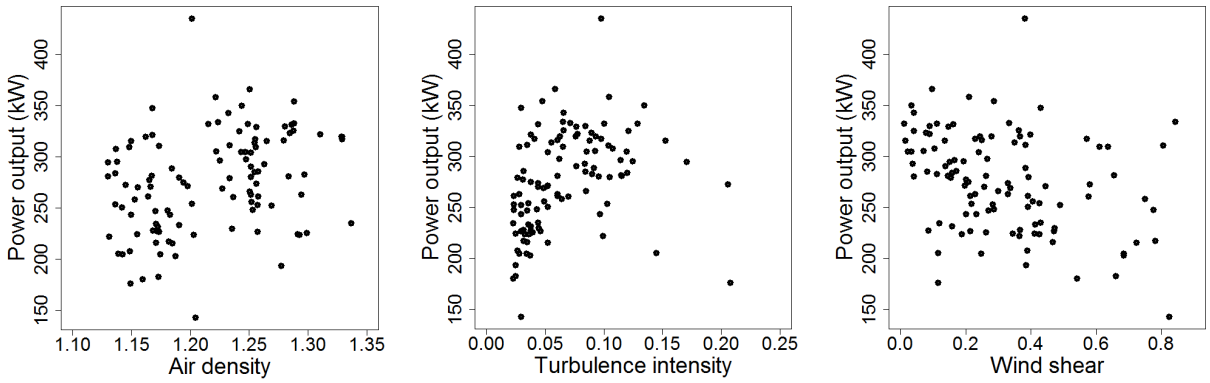
$$w_i(\boldsymbol{x}) = \frac{\mathcal{K}_{\boldsymbol{h}_x}(||\boldsymbol{x} - \boldsymbol{x}_i||)}{\sum_{i=1}^{N} \mathcal{K}_{\boldsymbol{h}_x}(||\boldsymbol{x} - \boldsymbol{x}_i||)}, \tag{5}$$

$\boldsymbol{h}_x = (h_1, \ldots, h_q)$ and h_y are bandwidth parameters controlling the smoothness in, respectively, the environmental factors, $\boldsymbol{x}$, and the power output, y, and q is the number of explanatory variables in $\boldsymbol{x}$. In our study, $q = 7$ for $\mathcal{D}_{\text{OSWF}}$ and $q = 5$ for $\mathcal{D}_{\text{ILWF}}$.

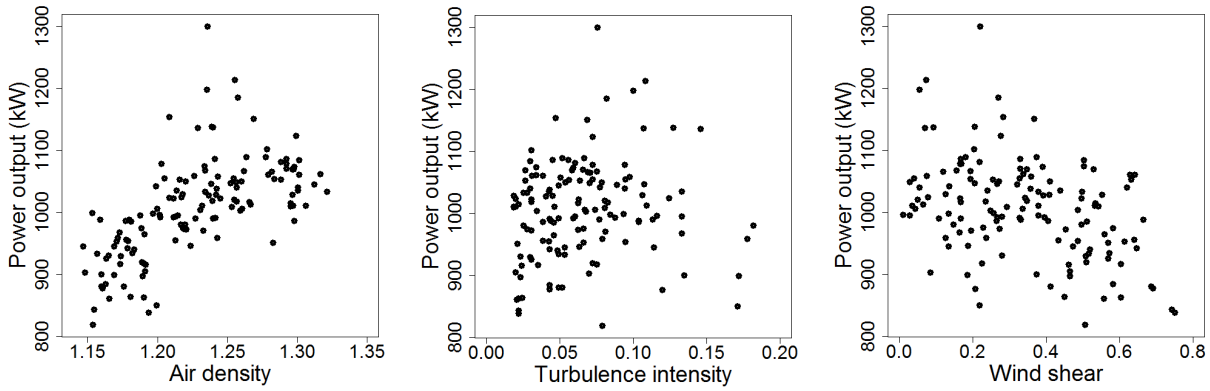
The above formulation contains kernel functions of two different dimensions, $\mathcal{K}_{h_y}(l)$ and $\mathcal{K}_{\boldsymbol{h}_x}(||\boldsymbol{l}||)$. $\mathcal{K}_{h_y}$ is a scaled kernel function and takes the form of $h_y^{-1} K(\frac{l}{h_y})$, where $K(\cdot)$ is assumed to be a real valued, integrable and non-negative even function. In our study, K is chosen to be a univariate Gaussian kernel function. $\mathcal{K}_{\boldsymbol{h}_x}(||\boldsymbol{l}||)$ is a multivariate kernel function and is composed of a product kernel that is a multiplication of univariate kernel functions, such as

$$\mathcal{K}_{\boldsymbol{h}_x}(||\boldsymbol{l}||) := \mathcal{K}_{h_1}(l_1) \mathcal{K}_{h_2}(l_2) \cdots \mathcal{K}_{h_q}(l_q), \tag{6}$$

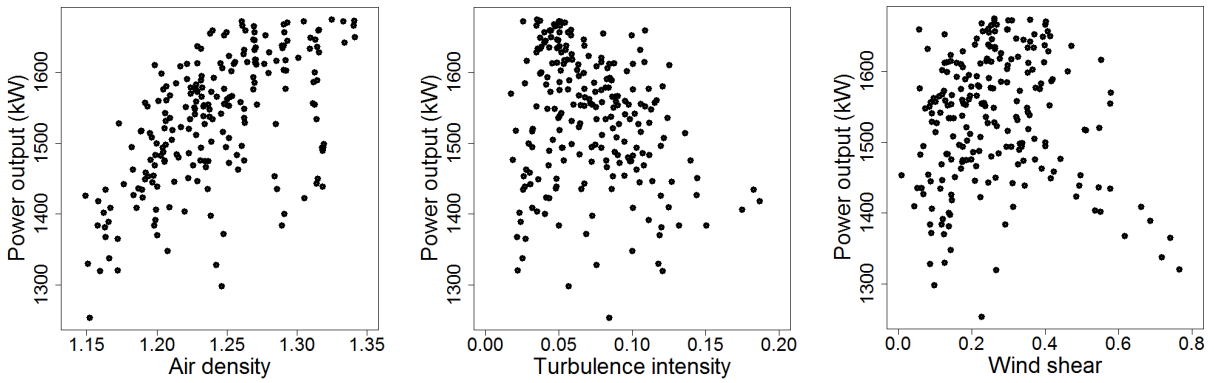
where $\mathcal{K}_{h_j}(l_j)$ is generally a univariate Gaussian kernel as well, except for wind direction, D. The kernel function for D is chosen to be the von Mises kernel (Taylor 2008), because it is a circular variable that may cause trouble in numerical computation; for more comprehensive discussion regarding the handling of circular variables, please refer to (Marzio et al. 2012; Marzio et al. 2013; Marzio et al. 2014). The von Mises kernel function can characterize the



(a) Air density (b) Turbulence intensity (c) Wind shear



(d) Air density (e) Turbulence intensity (f) Wind shear



(g) Air density (h) Turbulence intensity (i) Wind shear

Figure 4: Scatter plots of the power output versus environmental factors under specific wind speeds and wind directions. Top Panel: $6.1 < V < 6.2, 270 < D < 300$; Middle Panel: $9.1 < V < 9.2, 270 < D < 300$; and Bottom Panel: $11.1 < V < 11.2, 270 < D < 300$.

directionality of a circular variable and takes the form

$$\mathcal{K}_\nu(D - D_i) = \frac{\exp\{\nu \cos(D - D_i)\}}{2\pi I_0(\nu)} \tag{7}$$

where $I_0(\cdot)$ is the modified Bessel function of order 0, and ν is the concentration parameter of the von Mises kernel, which has now taken the role of the inverse of the bandwidth parameter h_2.

In addition, the mean of the conditional density estimator in (4) provides an estimator of the conditional mean function, $m(\boldsymbol{x}) := E(y|\boldsymbol{x})$, as

$$\hat{m}(\boldsymbol{x}) = \int y\hat{f}(y|\boldsymbol{x})dy. \tag{8}$$

Hyndman et al. (1996) noted that the estimator in (8) is equivalent to the Nadaraya-Watson (NW) regression estimator and only depends on $\boldsymbol{h}_x$, the smoothing parameter related to $\boldsymbol{x}$. The Nadaraya-Watson estimator is

$$\hat{m}(\boldsymbol{x}) = \sum_{i=1}^{N} w_i(\boldsymbol{x})y_i. \tag{9}$$

In the remainder of the paper, we will use the expression in (9) as our mean function estimator.

In the current dataset, the environmental factors are five to seven. As the wind industry currently arranges the data in 10-minute blocks, one year worth of data translates to slightly over 52,000 data points, which can still become scarce in a seven-dimensional factor space. This is to say, once the data points are dispersed into the seven-dimensional space, certain combinations of environmental conditions could have very little data or even no data at all, and therefore deteriorates the performance of the resulting multivariate kernel model. If a technology innovation makes additional measurements available so that the model could

entertain more than seven explanatory variables, the current CKD approach would run into a scalability problem. Running a high-dimensional CKD will also take longer computation times than practitioners typically prefer. It is thus desirable to use fewer input variables to form the multivariate product kernels if possible.

Our tailored power curve modeling is to devise an additive multivariate kernel model. Let us present the mathematical expression of the kernel first and then elaborate its merit.

For notation simplicity, we designate the first two elements of $\boldsymbol{x}$ of both $\mathcal{D}_{\text{ILWF}}$ and $\mathcal{D}_{\text{OSWF}}$, namely x_1 and x_2, as V and D, respectively. We introduce a new symbol, $\boldsymbol{x}^j$, $j = 3, ..., q$, such that $\boldsymbol{x}^j := (x_1, x_2, x_j)$. With this notation, we propose to estimate the density of y, conditional on an $\boldsymbol{x}$, by using

$$\hat{f}(y|\boldsymbol{x}) = \sum_{i=1}^{N} \frac{1}{(q-2)} \left[w_i(\boldsymbol{x}^3) + \ldots + w_i(\boldsymbol{x}^q) \right] \mathcal{K}_{h_y}(y - y_i), \tag{10}$$

and the conditional mean function by

$$\hat{m}(\boldsymbol{x}) = \frac{1}{(q-2)} \left[\hat{m}(\boldsymbol{x}^3) + \ldots + \hat{m}(\boldsymbol{x}^q) \right]. \tag{11}$$

As in the above expression, our resulting model keeps the multivariate kernels but we limit them to be product kernels of three inputs. Based on the observations from Figures 3 and 4, we decide to include V and D, incorporating wind speed and direction information, in every multivariate kernel so that the three-variable kernel can capture the interaction effect between the third environmental factor with wind speed and wind direction. Then, all the multivariate kernels constitute an additive model such that the resulting model has good scalability. The resulting model could be used for high-dimensional data without causing computational or data sparsity problems. When additional explanatory variables become available, we would envision to add extra additive terms, each of which has the same structure as the current terms, namely a three-variable multivariate kernel having inputs of

V, D, and a third explanatory variable.

3.3 Bandwidth selection

The key parameters in our kernel model are the bandwidths h_y and $\boldsymbol{h}_x$. In this study, we employ a data-driven selection criterion proposed by Hall, Racine and Li (2004) and Fan and Yim (2004), known as the integrated squared error (ISE) criterion, as follows:

$$\begin{aligned} \mathrm{ISE}(\boldsymbol{h}_x, h_y) &= \int\int \Big(f(y|\boldsymbol{x}) - \hat{f}(y|\boldsymbol{x})\Big)^2 f(\boldsymbol{x}) dy d\boldsymbol{x} \\ &= \int\int \hat{f}(y|\boldsymbol{x})^2 f(\boldsymbol{x}) dy d\boldsymbol{x} - 2\int\int \hat{f}(y|\boldsymbol{x}) f(y|\boldsymbol{x}) f(\boldsymbol{x}) dy d\boldsymbol{x} \\ &\qquad + \int\int f(y|\boldsymbol{x})^2 f(\boldsymbol{x}) dy d\boldsymbol{x} \\ &= I_1 - 2I_2 + I_3. \end{aligned} \tag{12}$$

With this criterion, one would choose the bandwidths that minimize the ISE. Because I_3 in the ISE expression does not depend on the bandwidth selection, it can be omitted during the minimization of ISE.

Fan and Yim (2004) suggested leave-one-out cross-validation estimators of I_1 and I_2 as

$$\begin{aligned} \hat{I}_1 &= \frac{1}{N}\sum_{i=1}^{N} \int \Big(\hat{f}_{-i}(y|\boldsymbol{x}_i)\Big)^2 dy, \text{ and} \\ \hat{I}_2 &= \frac{1}{N}\sum_{i=1}^{N} \hat{f}_{-i}(y_i|\boldsymbol{x}_i), \end{aligned} \tag{13}$$

where $\hat{f}_{-i}(y|\boldsymbol{x}_i)$ is the estimator $\hat{f}(y|\boldsymbol{x}_i)$ with observation i omitted. Practically, the data-driven bandwidth selection is simply to choose the bandwidths $\boldsymbol{h}_x$ and h_y that minimize $\hat{I}_1 - 2\hat{I}_2$.

Using this cross-validation algorithm could, however, take a long computational time; for the models and datasets we have at hand, the computation ran for more than a day before we manually stopped it. In order to have a faster bandwidth selection for practical purposes,

we choose to employ a much simpler, greedy procedure to select the bandwidth parameters one at a time, as described below.

- *Algorithm I*: Greedy kernel bandwidth selection

    1. Consider only a simple univariate kernel regression corresponding to individual environmental variables;
    2. Calculate the bandwidth for each univariate kernel following the direct plug-in (DPI) approach suggested by Ruppert, Sheather and Wand (1995). This DPI approach provides an optimal bandwidth formula, expressed below, that is supposed to minimize the asymptotically weighted integrated squared error:

$$\hat{h} = \left(\frac{1}{2\sqrt{\pi}}\right)^{1/5} \left[\frac{\hat{\sigma}^2(b-a)}{N\hat{\theta}_{22}^{0.05}}\right]^{1/5}, \tag{14}$$

    where $[a, b]$ is the range of each environmental variable, and $\hat{\sigma}^2$ and $\hat{\theta}_{22}$ are estimated from the data using the DPI algorithm; for details, please refer to Ruppert et al. (1995);
    3. Denote the resulting bandwidths as $(\hat{h}_1, \hat{h}_2, \ldots, \hat{h}_q)$;
    4. Use the most basic power curve model that includes only the wind speed, V and wind direction, D as inputs, and fix the bandwidths for the two univariate kernels corresponding to V and D as $\hat{h}_1$ and $\hat{h}_2$, respectively. Then, estimate the bandwidth $\hat{h}_y$ that minimizes $\hat{I}_1 - 2\hat{I}_2$.

In the above algorithm, in handling the von Mise kernel, we follow an approach suggested by Taylor (2008) that ties the concentration parameter ν to the bandwidth parameter h_2 as $\nu = \frac{1}{h_2^2}$. Then, h_2 can be selected together with other bandwidth parameters for the Gaussian kernels, as explained above.

As a greedy procedure, the algorithm cannot guarantee the optimality of the chosen bandwidths. But, as we will show in the subsequent section, our kernel model with the heuristically chosen bandwidths is able to produce remarkable error reduction as compared with the current power curve methods.

4. RESULTS

In this section, we evaluate the performance of our kernel-based approach using the wind farm measurements in $\mathcal{D}_{\text{ILWF}}$ and $\mathcal{D}_{\text{OSWF}}$ and compare its performance with the existing methods.

4.1 Performance criteria

We evaluate the performance of our method in terms of point estimation as well as density estimation. We therefore use two criteria: for point estimation, we use the root mean square error (RMSE), and for density estimation, we use the mean continuous ranked probability score (CRPS) (Gneiting and Raftery 2007). We randomly divide each data set into a partition of 80% for training and 20% for testing, and use the test dataset to make an out-of-sample evaluation of the above two criteria. Specifically, RMSE is computed as

$$\text{RMSE} = \sqrt{\frac{1}{N_{TS}} \sum_{i=1}^{N_{TS}} (\hat{m}(\boldsymbol{x}_i) - y_i))^2}, \tag{15}$$

where N_{TS} is the number of data points in a test data set. The CRPS compares the estimated cumulative distribution function (CDF) with the observed value. It is computed as

$$\text{CRPS} = \frac{1}{N_{TS}} \sum_{i=1}^{N_{TS}} \int \left(\hat{F}(y|\boldsymbol{x}_i) - \mathbb{1}(y > y_i) \right)^2 dy, \tag{16}$$

where $\hat{F}(y|\boldsymbol{x}_i)$ is the estimated CDF, given a setting of the environmental variable, $\boldsymbol{x}_i$, and $\mathbb{1}(\cdot)$ is the indicator function.

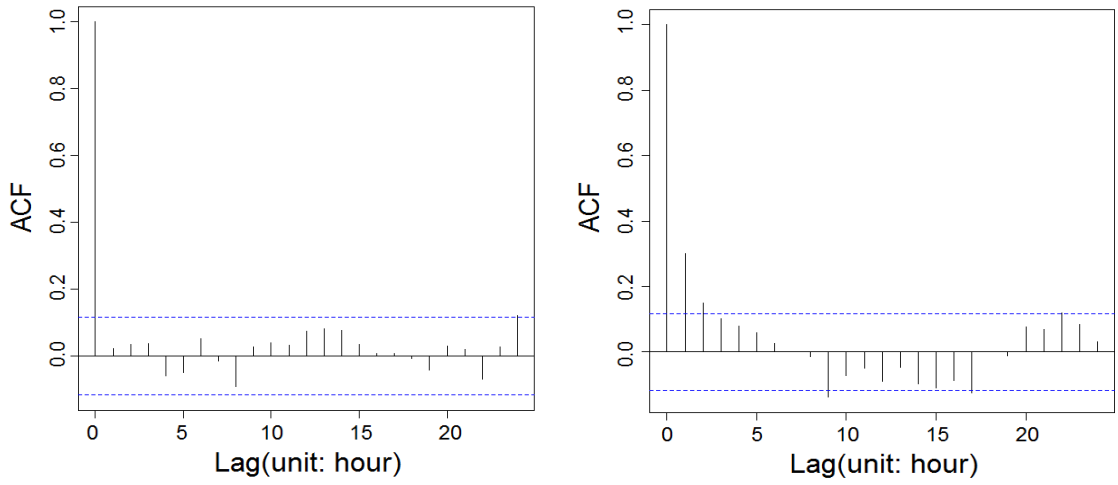
Algorithm I works well for all datasets for bandwidth selection. For point estimation, we are able to use all the training data for bandwidth selection and the computational time is of no concern at all. But for density estimation, even with the greedy algorithm, the last step (Step 4 above) that finds the bandwidth for y still takes a long running time, had we used all the training data. In the end, we decide to randomly select 25% of the training data for bandwidth selection in density estimation.

For the out-of-sample testing purpose, we are able to use all the testing data points for computing the out-of-sample RMSE values. But for computing the CRPS values, using all the testing data again requires more than ten hours of computation; it is feasible but not always practical. We experiment with a randomly sampled subset of 1,000 data points from the test set and find that using 1,000 data points to calculate the CRPS values remains reasonably stable over different random sampling. So, in the subsequent sections, we report the CRPS values based on 1,000 testing data points.

A question may arise as of why we use a random split of training/test data to measure performance. Recall that the focus of this paper is to estimate a functional relationship between y and $\boldsymbol{x}$, say $f(\cdot)$. The $f(\cdot)$ is fundamentally decided by a turbine's own aerodynamic characteristics, primarily through its own design and manufacturing, and can be reasonably assumed unchanged over a short period of time, say six months or a year. Our objective is to see how well a set of $(\boldsymbol{x}, y)$ data pairs, collected in several months, can help recover this $f(\cdot)$, presumably stationary over the same period. One important consideration that we use random splitting is to ensure that the training data and the test data represent the same spectrum of the weather conditions $\boldsymbol{x}$. We believe this is a fair approach to test our model.

We arrange the data in random order as well as in time series order, producing two sets of models, and then compute the autocorrelation function of the respective model residuals produced from the test data. Figure 5 presents the autocorrelation function plots. As expected, the autocorrelation in the residuals appear stronger when data are arranged in time

series but still the absolute correlation level is low. More importantly, the autocorrelation drops to a nearly negligible level (below 0.2) after one hour. Considering that our test data are of time durations ranging from a couple of weeks to a couple months, this slight correlation among the first hour does not appear to present a problem.



(a) Random order (b) Time series order

Figure 5: Autocorrelation function values after an AMK model fit, using an inland turbine dataset. The physical unit of lag (x-axis) is hour.

4.2 Important environmental factors affecting power output

From the physical understanding presented in Section 3.1, we believe that wind speed, direction, and air density should be important factors to be included in a power curve model. The question is what else may also need to be included. This section sets out to find what set of environmental factors makes the best prediction for a given data set. In the following, we present the results of using point estimates and RMSE values.

Our first set of results is to show the RMSE values when the additive multivariate kernel model includes a single additive term from $\boldsymbol{x}^3$ to $\boldsymbol{x}^q$. Recall that each additive term is a three-variable multivariate kernel with two of the variables always being the wind speed, V

and wind direction, D.

We choose as the baseline model for comparison the kernel model that has only the wind speed and wind direction (V, D) in a product kernel. In fact, this bivariate kernel (BVK) model is the same as the one used by Jeon and Taylor (2012). The results are shown in Table 2.

Table 2: Impact on RMSE when including different environmental factors. The notation of $(\cdot, \cdot, \rho)$ means that the additive term included in the model has the wind speed, V and wind direction D, and air density, ρ, as its inputs, where the wind speed and wind direction are shorthanded as two dots. Other notations follow the same convention. The percentages in the parentheses are the reduction in terms of RMSE when the corresponding model's point estimation is compared with that of BVK.

WT	BVK	$(\cdot, \cdot, \rho)$	$(\cdot, \cdot, I)$	$(\cdot, \cdot, S_b)$	$(\cdot, \cdot, S_a)$	$(\cdot, \cdot, H)$
WT1	148.2	126.4	144.8	145.9	·	·
		(14.7%)	(2.3%)	(1.6%)	·	·
WT2	154.7	136.6	150.3	152.1	·	·
		(11.7%)	(2.8%)	(1.7%)	·	·
WT3	144.5	118.2	137.0	131.8	·	·
		(18.2%)	(5.2%)	(8.8%)	·	·
WT4	209.4	179.7	192.6	196.7	·	·
		(14.2%)	(8.0%)	(6.1%)	·	·
WT5	270.8	245.2	275.4	294.0	268.6	257.0
		(9.5%)	(-1.7%)	(-8.6%)	(0.8%)	(5.1%)
WT6	291.8	249.3	290.1	285.9	280.6	264.7
		(14.6%)	(0.6%)	(2.0%)	(3.8%)	(9.3%)

Based on these results, we make the following observations:

- In both the inland wind farm and offshore wind farm, air density, ρ, is indeed, after the wind speed and wind direction, the most significant factor in wind power generation. Including ρ in the model delivers reductions in RMSE from 10% to 18% across the board. This outcome is consistent with the physical understanding expressed earlier.

- For the offshore wind turbines, humidity, H, appears to be another important factor in

explaining variations in power outputs. Unfortunately, we will not be able to know for certain whether humidity is also a significant factor in the inland wind farm because its measurements were not available in our dataset. Given its significance in the offshore farm, this should provide strong enough motivation for practitioners to measure humidity at some inland wind farms and test the hypothesis.

- The remaining three factors, namely turbulence intensity and the two wind shears, which each represents some other aspects of wind dynamics, show also mostly positive impact, except for the case of WT5. These wind dynamics effects appear to be more pronounced for the inland turbines than the offshore ones. The numerical analysis indicates that the significance of these effects are after that of ρ and H.

The next step we undertake is to determine which other factors may impact the power output when we include more than one additive term in our model, conditional on the factors that have already been included. Based on the observations expressed above, for both inland and offshore turbines, the first additive term included is always (V, D, ρ). For the inland turbines, in addition to this first term, there are two more terms that have either turbulence intensity, I, or the below-hub wind shear, S_b. For the offshore turbines, we also always include a second additive term (V, D, H). Then, in addition to the first two terms, there are three more terms that have either the two wind shears, S_a, S_b, or turbulence intensity, I. The two wind shears are always included or excluded together in the numerical analysis to keep the total number of model comparisons manageable. Tables 3 and 4 present the model comparison results.

For some of the inland turbines, the best additive multivariate kernel model explaining their power output includes the input factors of the wind speed and wind direction (V and D), air density (ρ), and turbulence intensity (I), while some others include the wind speed and wind direction (V and D), air density (ρ), and wind shear (S_b). These versions differ marginally. In the next subsection where the additive multiplicative kernel model is

Table 3: Model comparisons using data in $\mathcal{D}_{\text{ILWF}}$. RMSE values are reported in the table. Boldface values are the smallest RSME in the row.

WT	$(\cdot, \cdot, \rho)$	$(\cdot, \cdot, \rho, I)$	$(\cdot, \cdot, \rho, S_b)$	$(\cdot, \cdot, \rho, I, S_b)$
WT1	126.4	**125.8**	125.9	126.9
WT2	136.6	**136.0**	136.9	136.6
WT3	118.2	115.1	**111.5**	112.7
WT4	179.7	**174.8**	177.3	176.2

Table 4: Model comparisons using data in $\mathcal{D}_{\text{OSWF}}$. RMSE values are reported in the table. Boldface values are the smallest RSME in the row.

WT	$(\cdot, \cdot, \rho, H)$	$(\cdot, \cdot, \rho, H, I)$	$(\cdot, \cdot, \rho, H, S_a, S_b)$	$(\cdot, \cdot, \rho, H, I, S_a, S_b)$
WT5	**236.1**	239.2	244.1	245.1
WT6	**242.5**	248.7	250.8	254.8

compared with other methods, we choose the model with four factors, V, D, ρ, and I, as the "best model" for the inland turbines, because when it does not produce the smallest RMSE (only one case), the difference between its RMSE and the smallest RMSE is around three percentage points.

For the offshore turbines, it is rather clear that the model with the wind speed(V), wind direction (D), air density (ρ), and humidity (H) produces the lowest RMSE. Including other environmental factors in the model could instead increase the RMSE. The increase in RMSE is consistent and can be as much as 5.1% for turbine WT6. In the next section, we choose the model with V, D, ρ, and H as the "best model" for the offshore turbines.

If we repeat the above analysis using the CRPS measure, the insights remain the same. We therefore omit the presentation of the detailed CRPS results.

4.3 Comparison of the estimation accuracy of different models

In this subsection, we compare the "best" additive multivariate kernel model, selected in the preceding subsection, with four different methods: the binning method (BIN), popular in the wind industry and arguably the most widely used method in practice, BVK (Jeon

and Taylor 2012), BART (Chipman et al. 2010), and SSANOVA (Gu 2013). Our proposed additive multiplicative kernel method is represented by AMK. Recall that the binning method we use here is the version having incorporated the air density adjustment. To make this explicit, we use the notation BIN_a. Note that the RMSE values of BVK and AMK in this section differ slightly from those in Tables 3 and 4 because we split the training and test datasets randomly, so that the training/test datasets used in this subsection are not exactly the same as those in the previous section.

The comparison results using the inland turbines are included in Table 5 and Table 6. These two tables show the RMSE-based comparison and the CRPS-based comparison, respectively. Note that the binning method can produce only point estimation, while BVK, BART and AMK produces both point and density estimations. SSANOVA is supposed to produce density estimation as well, but doing so takes way too long time; we have to manually stop it after 15 hours computation. So, in the CRPS-based comparison, only are BVK, AMK and BART included. In the RMSE-based comparisons, the baseline model used in the table is the binning method, and in the CRPS-based comparison, the baseline model is BVK. In the tables, the lowest values of RMSE or CRPS are highlighted in boldface fonts.

Table 5: Comparing RMSE using data from $\mathcal{D}_{\text{ILWF}}$

Turbine	BIN_a	BVK	AMK		BART		SSANOVA	
			$(\cdot,\cdot,\rho)$	$(\cdot,\cdot,\rho,I)$	$(\cdot,\cdot,\rho)$	$(\cdot,\cdot,\rho,I)$	$(\cdot,\cdot,\rho)$	$(\cdot,\cdot,\rho,I)$
WT1	220.4	146.9	**123.2**	126.1	131.5	127.8	149.3	142.4
		(33.3%)	(44.1%)	(42.8%)	(40.3%)	(42.0%)	(32.3%)	(35.4%)
WT2	201.6	149.6	**131.2**	132.2	137.3	133.3	147.9	153.8
		(25.8%)	(34.9%)	(34.4%)	(31.9%)	(33.9%)	(26.6%)	(23.7%)
WT3	219.2	149.7	**119.4**	119.6	129.1	120.7	149.0	142.1
		(31.7%)	(45.5%)	(45.4%)	(41.1%)	(44.9%)	(32.0%)	(35.2%)
WT4	265.7	193.9	172.1	**168.1**	187.6	176.9	199.7	190.8
		(27.0%)	(35.2%)	(36.7%)	(29.4%)	(33.4%)	(24.8%)	(28.2%)

Table 6: Comparing CRPS using data from $\mathcal{D}_{\text{ILWF}}$

Turbine	BVK	AMK		BART	
		$(\cdot,\cdot,\rho)$	$(\cdot,\cdot,\rho,I)$	$(\cdot,\cdot,\rho)$	$(\cdot,\cdot,\rho,I)$
WT1	84.0	77.3 (8.0%)	75.5 (10.2%)	74.9 (10.9%)	**72.6** (13.6%)
WT2	85.5	79.3 (7.3%)	**77.2** (9.7%)	87.9 (-2.7%)	80.8 (5.5%)
WT3	84.6	**69.1** (18.3%)	69.2 (18.3%)	78.9 (6.8%)	70.2 (17.1%)
WT4	114.9	**96.3** (16.2%)	96.5 (16.0%)	128.3 (-11.7%)	117.9 (-2.6%)

We notice that the BVK model produces a significant improvement over the industry standard binning method, with a reduction of RMSE ranging from 26% to almost 33%. Our additive multivariate kernel method improves further from BVK another 10% to 14%. In other words, the additive multivariate kernel method can reduce RMSE from the binning method by 35% to 45%. We believe that the improvement is the result of including the additional environmental factors in the model. When compared with BART and SSANOVA, AMK produces similar, but slightly better, results than BART in terms of point estimation, while the RMSE from the SSANOVA method is not as competitive as AMK and BART. In fact, SSANOVA performs much closer to what BVK does.

In Table 6, when using data from WT1 through WT3, we notice that both AMK and BART produce better density estimation than BVK, whereas they two perform comparably. In the case of WT4 data, BART performed noticeably worse than AMK, and even BVK. We observe that the WT4 data have the largest variations among the four inland turbine datasets, as evident by its large RMSE/CRPS values. This large variation could be due to the fact that WT4 is located the farthest away from its companion mast so that the wind measurements taken at the mast are less representative of the wind condition at the turbine

site. This case appears to suggest the sensitivity of the BART model to the data quality.

Tables 7 and 8 present the comparison results for the offshore turbines. In terms of point estimation, AMK is 8% to 12% better than BART, 11% to 14% better than SSANOVA, 10% to 14% better than BVK, 16% to 25% better than the binning method, whereas BART, BVK and SSANOVA perform comparably. In terms of density estimation, AMK is about 15% better than BVK, whereas BART is 15% to 20% worse than BVK, or equivalently, AMK is 24% to 30% better than BART. The results using the offshore turbines data appear to produce a similar message as in the case of inland turbine WT4: they all have a high level of noise, and in all three cases, BART appears to suffer from the elevated level of noise in data.

Table 7: Comparing RMSE using data from $\mathcal{D}_{\text{OSWF}}$

Turbine	BIN_a	BVK	AMK		BART		SSANOVA	
			$(\cdot,\cdot,\rho)$	$(\cdot,\cdot,\rho,H)$	$(\cdot,\cdot,\rho)$	$(\cdot,\cdot,\rho,H)$	$(\cdot,\cdot,\rho)$	$(\cdot,\cdot,\rho,H)$
WT5	302.7	281.7	255.7	**254.5**	278.1	275.7	285.7	289.9
		(6.9%)	(15.5%)	(15.9%)	(8.1%)	(8.9%)	(5.6%)	(4.2%)
WT6	328.4	284.1	**245.0**	248.1	282.2	279.1	293.2	285.3
		(13.5%)	(25.4%)	(24.5%)	(14.1%)	(15.0%)	(10.7%)	(13.1%)

Table 8: Comparing CRPS using data from $\mathcal{D}_{\text{OSWF}}$

Turbine	BVK	AMK		BART	
		$(\cdot,\cdot,\rho)$	$(\cdot,\cdot,\rho,H)$	$(\cdot,\cdot,\rho)$	$(\cdot,\cdot,\rho,H)$
WT5	130.9	116.2	**111.5**	152.5	151.5
		(11.2%)	(14.8%)	(-16.6%)	(-15.6%)
WT6	146.1	124.1	**122.1**	177.9	176.3
		(15.1%)	(16.5%)	(-21.8%)	(-20.6%)

5. CONCLUDING REMARKS

This study presents an additive multivariate kernel method for modeling power curves with a variety of environmental factors. It is an appealing approach because this new power curve model can capture the nonlinear relationships between environmental factors and the wind power output, as well as the high-order interaction effects among some of the environmental factors. The new power curve model generally outperforms its competitors in terms of prediction errors for both point estimation and density estimation. We believe that AMK's good performance does not happen by chance. However simple, our method makes explicit use of the physical understanding behind wind energy production for devising the tailored kernel model structure, whereas in BART or SSANOVA, by contrast, the intrinsic structure is left to be learned through the data. SSANOVA does not seem capable enough to capture the structure as much as the tailored AMK does. BART is more capable than SSANOVA but its capability is still worse than the human-guided physics-based understanding, and its capability becomes noticeably less effective when data are noisy.

In addition to its general good performance, we believe that the merit of our model lies in two additional aspects. The first one is its simplicity, which is a virtue valued by industrial practitioners, who generally do not like to blindly use a black-box method of which they do not understand well how the information fed into it was manipulated.

Another advantage is its fast computation for producing a point estimation. When calculating a mean power curve, AMK takes up to a few minutes, depending on the dataset sizes, while BART takes up to 40 minutes (because BART always produces distributions) and SSANOVA, being the slowest, takes up to two hours. In other words, AMK can be 10 times faster than BART and 30 times faster than SSANONA. This level of computation times may not be a point of concern among academic researchers but it is certainly meaningful and relevant in practice. Although density estimation offers a fuller picture, computing the conditional means is a much more common exercise that practitioners need to do on a

routinely basis and do so for a large number of turbines (in the order of thousands for a major wind company). Being able to shorten a routine computation by an order of magnitude is definitely appreciated by practitioners.

REFERENCES

Ackermann, T. (2005), *Wind Power in Power Systems*, Hoboken, New Jersey: John Wiley & Sons.

Albers, A., Klug, H., and Westermann, D. (1999), "Power performance verification," in *Proceedings of European Wind Energy Conference*, Nice, France: March 1-5, pp. 657–660.

AWEA (2012), "U.S. wind industry fourth quarter 2011 market report," Technical report, American Wind Energy Association (AWEA). Available at http://www.awea.org/learnabout/industry_stats/upload/4Q-2011-AWEA-Public-Market-Report_1-31.pdf.

Belghazi, O., and Cherkaoui, M. (2012), "Pitch angle control for variable speed wind turbines using genetic algorithm controller," *Journal of Theoretical and Applied Information Technology*, 39, 5–10.

Chipman, H., George, E., and McCulloch, R. (2010), "BART: Bayesian additive regression trees," *The Annals of Applied Statistics*, 4, 266–298.

DOE (2008), "20% wind energy by 2030," Technical report, U.S. Department of Energy (DOE). Available at http://www.20percentwind.org/.

Fan, J., and Yim, T. (2004), "A cross-validation method for estimating conditional densities," *Biometrika*, 91, 819–834.

Giebel, G., Brownsword, R., Kariniotakis, G., Denhard, M., and Draxl, C. (2011), "The state-of-the-art in short-term prediction of wind power: A literature overview," Technical report, Risø National Laboratory, Roskilde, Denmark. Available at http://www.anemos-plus.eu/images/pubs/deliverables/aplus.deliverable_d1.2.stp_sota_v1.1.pdf.

Gneiting, T., Larson, K., Westrick, K., Genton, M. G., and Aldrich, E. (2006), "Calibrated probabilistic forecasting at the Stateline wind energy center: The regime-switching space-time method," *Journal of the American Statistical Association*, 101, 968–979.

Gneiting, T., and Raftery, A. (2007), "Strictly proper scoring rules, prediction, and estimation," *Journal of the American Statistical Association*, 102, 359–378.

Gu, C. (2013), *Smoothing Spline ANOVA*, New York: Springer-Verlag.

Hall, P., Racine, J., and Li, Q. (2004), "Cross-validation and the estimation of conditional probability," *Journal of the American Statistical Association*, 99, 154–163.

Hayes, B., Ilie, I., Porpodas, A., Djokic, S., and Chicco, G. (2011), "Equivalent power curve model of a wind farm based on field measurement data," in *Proceedings of IEEE PowerTech*, Trondheim, Norway: June 19-23, pp. 1–7.

Hering, A. S., and Genton, M. G. (2010), "Powering up with space-time wind forecasting," *Journal of the American Statistical Association*, 105, 92–104.

Hyndman, B., Bashtannyk, D., and Grunwald, G. (1996), "Estimating and visualizing conditional densities," *Journal of Computational and Graphical Statistics*, 5, 315–336.

IEC (2005), *IEC 61400-12-1 Ed 1, Wind Turbines-Part 12-1: Power Performance Measurements of Electricity Producing Wind Turbines*, Geneva, Switzerland: International Electrotechnical Commission.

Jeon, J., and Taylor, J. (2012), "Using conditional kernel density estimation for wind power density forecasting," *Journal of the American Statistical Association*, 107, 66–79.

Kusiak, A., Zheng, H., and Song, Z. (2009), "On-line monitoring of power curves," *Renewable Energy*, 34, 1487–1493.

Marzio, M. D., Panzera, A., and Taylor, C. C. (2012), "Smooth estimation of circular cumulative distribution functions and quantiles," *Journal of Nonparametric Statistics*, 24, 935 – 949.

Marzio, M. D., Panzera, A., and Taylor, C. C. (2013), "Nonparametric regression for circular responses," *Scandinavian Journal of Statistics*, 40, 238 – 255.

Marzio, M. D., Panzera, A., and Taylor, C. C. (2014), "Nonparametric regression for spherical data," *Journal of the American Statistical Association*, 109, 748 – 763.

Monteiro, C., Bessa, R., Miranda, V., Botterud, A., Wang, J., and Conzelmann, G. (2009), "Wind power forecasting: state-of-the-art 2009," Technical Report ANL/DIS-10-1, Argonne National Laboratory, Illinois, U.S. Available at http://www.osti.gov/energycitations/product.biblio.jsp?osti_id=968212.

Nadaraya, E. (1964), "On estimating regression," *Theory of Probability and its Applications*, 9, 141–142.

Nielsen, T., Nielsen, H., and Madsen, H. (2002), "Prediction of wind power using time-varying coefficient functions," in *Proceedings of the 15'th IFAC World Congress on Automatic Control*, Barcelona, Spain: July 21-26.

Osadciw, L. A., Yan, Y., Ye, X., Benson, G., and White, E. (2010), "Wind turbine diagnostics based on power curve using particle swarm optimization," in *Wind Power Systems (Green Energy and Technology)*, eds. L. Wang, C. Singh, and A. Kusiak, Berlin: Springer-Verlag, pp. 151–165.

Pinson, P., Nielsen, H., Madsen, H., and Nielsen, T. (2008), "Local linear regression with adaptive orthogonal fitting for wind power application," *Statistics and Computing*, 18, 59–71.

Rehman, S., and Al-Abbadi, N. (2005), "Wind shear coefficients and their effect on energy production," *Energy Conversion and Management*, 46, 2578–2591.

Rosenblatt, M. (1969), "Conditional probability density and regression estimates," in *Multivariate Analysis II*, ed. P. Krishnaiah, New York: Academic Process, pp. 25–31.

Ruppert, D., Sheather, S., and Wand, M. (1995), "An effective bandwidth selector for local least squares regression," *Journal of the American Statistical Association*, 90, 1257–1270.

Sanchez, I. (2006), "Short-term prediction of wind energy production," *International Journal of Forecasting*, 22, 43–56.

Stephen, B., Galloway, S. J., McMillan, D., Hill, D. C., and Infield, D. G. (2011), "A copula model of wind turbine performance," *IEEE Transactions on Power Systems*, 26, 965–966.

Taylor, C. (2008), "Automatic bandwidth selection for circular density estimation," *Computational Statistics and Data Analysis*, 52, 3493–3500.

Tindal, A., Johnson, C., LeBlanc, M., Harman, K., Rareshide, E., and Graves, A. (2008), "Site-specific adjustments to wind turbine power curves," in *Proceedings of 2008 AWEA WindPower Conference*, Houston, TX: June 2-4, pp. 1–11.

Uluyol, O., Parthasarathy, G., Foslien, W., and Kim, K. (2011), "Power curve analytic for wind turbine performance monitoring and prognostics," *Annual Conference of the prognostics and Health Management Society*, 2, Publication Control Number 049.

Wan, Y., Ela, E., and Orwig, K. (2010), "Development of an equivalent wind plant power curve," Technical Report NREL/CP-550-48146, National Renewable Energy Laboratory. Available at http://www.nrel.gov/docs/fy10osti/48146.pdf.

Watson, G. (1964), "Smooth regression analysis," *Sankhyā: The Indian Journal of Statistics, Series A*, 26, 359–372.

Yan, Y., Osadciw, A., Benson, G., and White, E. (2009), "Inverse data transformation for change detection in wind turbine diagnostics," in *Proceedings of the 22nd IEEE Canadian Conference on Electrical and Computer Engineering*, St. John, Newfoundland, Canada: May 3-6, pp. 944–949.

Zhu, X., and Genton, M. G. (2012), "Short-term wind speed forecasting for power system operations," *International Statistical Review*, 80, 2–23.