Proceedings of ASME Turbo Expo 2015: Turbine Technical Conference and Exposition
GT2015
June 15 – 19, 2015, Montréal, Canada

GT2015-43286

DATA ANALYTICS METHODS FOR WIND ENERGY APPLICATIONS

Yu Ding *
Industrial & Systems Engineering
Texas A&M University
College Station, Texas, 77843
Email: yuding@iemail.tamu.edu

Jiong Tang
Mechanical Engineering
University of Connecticut
Storrs, CT, 06269
Email: jtang@engr.uconn.edu

Jianhua Z. Huang
Statistics Department
Texas A&M University
College Station, Texas, 77843
Email: jianhua@stat.tamu.edu

ABSTRACT

In the wind industry, it is important to assess a turbine systems response under different wind profiles. For instance, a wind-to-power relationship is crucial for wind power forecast, and a wind-to-stress relationship is important for selecting critical design parameters meeting the reliability requirement. Given the complexity involved in a turbine system, it is impossible to write a neat, analytical expression to underlie the above-mentioned relationships. Almost invariably does the wind industry resort to data driven methods for a solution, namely that wind data and the corresponding turbine response data (bending moments or power outputs) are used together to fit empirically the functional relationship of interest. This paper presents a couple of nonparametric data analytic methods relevant to wind energy applications with real life example for demonstration.

NOMENCLATURE

y: turbine response;
$\mathbf{x}$: weather covariates or environmental variables;
x, x_i, x_j: elements in $\mathbf{x}$;
q: number of elements in $\mathbf{x}$;
$p(\cdot)$: probability density function;
$p(\cdot|\cdot)$: conditional probability density function;
$f(\cdot)$: a generic function;
$K(\cdot,\cdot)$: kernel function;
$w(\cdot)$: weighting coefficients in kernel regression or density estimation;
μ: location parameter of a GEV distribution;
σ: scale parameter of a GEV distribution;
ξ: shape parameter of a GEV distribution;
l_T: extreme load level for a service of T years;
k_1, k_2: knots in a spline model;
N: number of $(\mathbf{x}, y)$ pairs in a dataset;
Q: number of additive terms in an AMK model;
λ: bandwidth of a kernel function;
$B_\ell(\cdot)$: basis function in MARS models;
β_ℓ: coefficients of the basis functions in MARS models;
L: number of the basis functions in a MARS model;
$\mathbb{C}_n^m$: the number of combinations when choosing m out of n.

INTRODUCTION

Wind energy plays an increasingly important role in energy sustainability of the nation. According to [1], wind power accounts for 4.31% of all generated electrical energy in the US by July 2014. Albeit still a small percentage in the absolute sense, the percentage was less than 0.5% a decade ago, translating to an annualized growth of 29%. But wind energy's further growth has a lot to do with its operation efficiency or cost competitiveness of wind power production.

There is a general class of data analytics problems that are of interest to wind turbine's operations. This class of problem is

*Address all correspondence to this author.

to attain the following distribution:

$$p(y) = \int p(y|\mathbf{x})p(\mathbf{x})d\mathbf{x}, \tag{1}$$

where y is a turbine response, including, for instance, a power output response, or a load (stress) response, or anything that may be of interest, and $\mathbf{x}$ is the vector of weather-related covariates or environmental variables. Potential elements in $\mathbf{x}$ are: wind speed, wind direction, air density, humidity, turbulence intensity, wind shears, among others.

Being a variable energy source, controlling $\mathbf{x}$ is impractical in wind power production. Then modeling the conditional density $p(y|\mathbf{x})$ is of particular interest to engineers who design and operate the wind energy systems, as it provides turbine-specific insights concerning how a turbine reacts under a given weather profile.

Our research team studied two types of applications in which y takes different meanings. One is when y is the power output of a turbine. Then $p(y|\mathbf{x})$ is the probability density function (pdf) of the power curve, or equivalently, the mean function of $p(y|\mathbf{x})$ is the power curve. The other is when y is the maximum edge-wise or flap-wise bending moment measured at the root of turbine blades. Then $p(y|\mathbf{x})$ measures a reliability aspect of turbine operations. Specifically $p(y|\mathbf{x})$ is called the short-term distribution in the extreme load analysis [2].

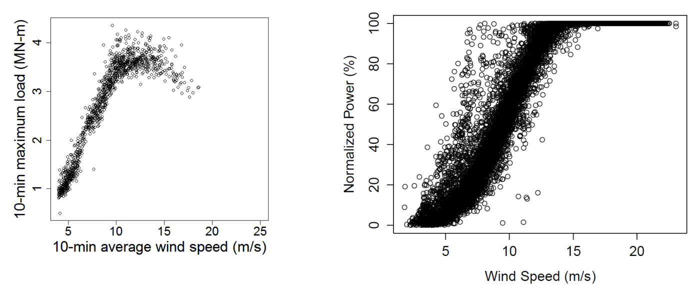
Modeling the conditional density $p(y|\mathbf{x})$ is beneficial to the wind energy industry. For instance, in the case of extreme load analysis (when y is the maximum bending moment), turbine manufacturers usually test a small number of representative turbines at their own testing site, producing $p(y|\mathbf{x})$. When a turbine is to be installed at a commercial wind farm, the weather profile at the proposed installation site can be collected and substituted into equation (1) as $p(\mathbf{x})$, so that the site-specific load response of the same type of turbine can be evaluated. In the case of power curve analysis (when y is the power output), the conditional modeling can control for the effect coming from the weather covariates or the environmental factors. Consequently, detecting changes in $p(y|\mathbf{x})$ can reveal a turbine's intrinsic change in its own aerodynamic properties and power production efficiency. We believe that the importance of the conditional density modeling is evident.

In fact, our research team has undertaken substantial research efforts and proposed sophisticated statistical methodologies in addressing the challenges in modeling the conditional density $p(y|\mathbf{x})$ in the context of wind energy applications [2–4]. However, these research reports, comprehensive in their undertaking, may not connect well with the practitioners, as they are written in statistical vernacular and immersed with statistical details and nicety. The purpose of this exposition is to provide an understanding why the conditional density problem is not trivial and then describe in engineers' language the data analytic methods that are potentially useful. Examples from wind energy applications are also presented to illustrate the impact that a sophisticated statistical method can make, as compared to the current industry practice.

The paper unfolds as following. The next section describes the technical challenges facing the conditional density modeling task and explains the current industrial practices. The subsequent section presents two data analytic methods commonly used in the statistical community: the spline method and the kernel method. The section following connects the generic data analytic methodologies with the wind energy applications, in which actual wind turbine data are used to exemplify the impact of the newly developed modeling techniques. Finally we summarize the paper.

TECHNICAL CHALLENGES AND INDUSTRIAL PRACTICE

There exists a number of technical challenges in modeling the conditional density $p(y|\mathbf{x})$. The first challenge is that the response data exhibits different types of nonlinearity and heteroscedasticity. Figure (1) presents examples of load and power output responses, plotted against wind speed. It is evident that both the mean and variance of the responses vary as the wind speed changes. Our past experiences allowed us to have observed other shapes of the response functions, unlike either shown in Figure (1), on turbines using different control mechanisms.



Stress-versus-wind data **Power-versus-wind data**

FIGURE 1. Nonlinearity and heteroscedasticity exhibited in turbine responses. In the wind-to-power data of the right panel, the power is normalized, due to a confidentiality agreement with the data provider.

The second challenge is the existence of multivariate factors in affecting a turbine's response. Although wind speed is arguably the most important factor, the effect of other environmental variables, such as wind direction or air density, cannot be ignored. For both the load response and the power response, there exists no analytical form of expression revealing how these factors affect the respective responses. It is known that certain factor interactions, in addition to the factors themselves, are also important but it is not yet fully understood of all the important factor interactions, let alone in which form the interactions take effect.

Due to these complexities and lack of simple parametric methods in modeling a turbine's response, the prevailing industry practice is to use an approach known as the binning method [5,6]. The idea of the binning method is simple: it first discretizes the domain of a weather vector, i.e., $\mathbf{x}$, into a finite number of bins. Within each bin, a stationary distribution is assumed for $p(y|\mathbf{x})$ with a parametric form. Then the parameters associated with this localized conditional distribution are estimated using the data falling into that bin. In the end, the $p(y|\mathbf{x})$ over the whole domain of $\mathbf{x}$ is the collection of $p(y|\mathbf{x})$'s estimated for each individual bin. Figure (2) illustrates the procedure of binning for producing a power curve. In doing so, the mean of $p(y|\mathbf{x})$ is estimated; but if needed, the data points in each bin can be used to estimate the distribution as well, should a parametric form (e.g., a beta distribution) be specified *a priori*.

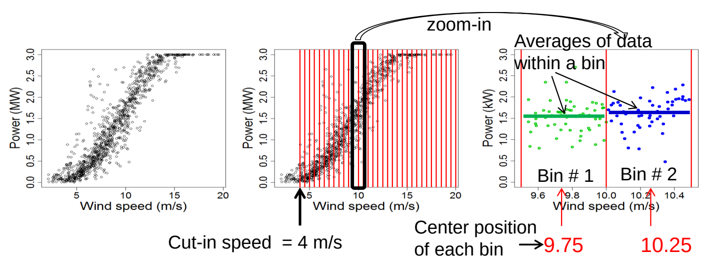


FIGURE 2. Binning method to produce a power curve using observed power output and wind speed data.

The use of the binning method is not without virtue. Its popularity is rooted in the fact that it provides a simple and easy-to-understand way of modeling a non-stationary and heteroscedastic probability distribution. It is nonparametric in nature and relying on few assumptions, implying its robustness in practice. It is capable of capturing local features in a turbine system's responses, which could vary across a range of weather conditions.

The major limitation in the use of binning method lies in its rigid compartmentalization of data and its separate use of data for individual bins, rather than borrowing strength from other bins. It becomes especially problematic when being extended to model multiple environmental factors, under which circumstance, using the binning method runs into the difficulty of "curse of dimensionality." Even a whole year worth of data could become scarce once they are dispersed into numerous bins under a moderate size of dimensions. We talk about up to ten dimensions here, which are not really high. But with ten dimensions, if each input variable is partitioned into 5 bins, the binning action produces nearly 10 million bins through variable combinations. Placing a single data point into each bin would require a data set of 10 million data records. The binning method is indeed a simple method, but it does not mean that the method produces a simple model in the end. In fact, a binning-based model can easily have several hundreds of parameters, overly complicated in the data-driven modeling practice and often suffering a poor performance due to this unwanted model complexity.

DATA ANALYTIC METHODS

The struggle with modeling the nonlinear responses like $p(y|\mathbf{x})$ in the wind energy applications can be placed in a broader context. Among the data-driven modeling approaches, the two ends of the methodology spectrum are the local approaches and the global approaches. Binning is a typical local method, as each of its modeling elements only capture what happens within a bin without worrying about anything else. Global methods refer to those that assume a single function form over the entire domain of the inputs, and the global function form can then be specified by a handful of parameters. A typical global model is the family of polynomial models.

Local and global models have their own pros and cons. Local models uses a lot of parameters so that it can be adaptive to local features, whereas the global models use far fewer parameters so its flexibility is limited. On global models, when one makes a change to the fitting outcome at one location, the fitting outcome at other places will be affected, as the whole model is constrained by the underlying function chosen *a priori*. Unlike local models, as we argue about the limitation of the binning method, the global models, especially the so-called additive models [7], are scalable, meaning that they can easily handle a large number of inputs without necessarily requiring a humongous amount of data. Figure (3) illustrates the spectrum of models.

In this section, we present two data analytic modeling approaches that fall in between the local and global approaches and can take advantage of both modeling philosophies. We focus more on the generic modeling approaches in this section and defer to the next section how they can be used in wind energy applications.

Spline methods

Spline methods in fact injects the idea of binning into the action of modeling a nonlinear response. Consider cubic splines [8] as example and see Figure 4 for illustration. A cubic spline partitions the input domain into a few segments (an action of binning) and models each segment using a cubic polynomial. In order to produce a smooth, coherent model for the whole domain, a cubic spline imposes continuity and smoothness constraints at the

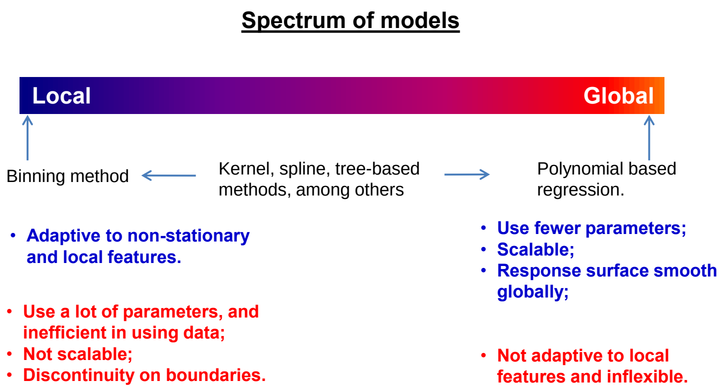


FIGURE 3. Struggle between local versus global approaches in data-driven modeling.

partition points; in Figure 4, the partition points are k_1 and k_2, known as knots. Knots do not have to be evenly spaced. Each cubic polynomial is specified by four parameters, producing a total of 12 parameters for the three piecewise cubic polynomials. The constraints imposed at the partition points, however, reduce the number of actual parameters that need to be estimated. For the cubic spline in Figure 4, there are three constraints at each partition point, requiring the equality of the function value, its first-order derivative and second-order derivative at the partition point. With the six constraints considered, the number of parameters for the cubic spline is six.

If only using the idea of binning without the boundary constraints, the response will look like the plot in the right-most panel. The three unconstrained piecewise cubic polynomials need a total of 12 parameters to specify. When a single global cubic polynomial is used to model a response, it uses four parameters, but its modeling adaptivity to local feature is far worse than the other two alternatives. With only a slight increase in model complexity (measured by the number of parameters), the cubic spline is endowed with the level of modeling adaptivity as a binning method allows.

People may argue that the binning method can use a single constant for each bin, so that the number of parameters for the right-most example in Figure 4 can be three, instead of 12. The problem of this argument is that when using a constant to model a bin, the bin width needs to be much smaller, or equivalently, the number of bins needs to be much greater, so that a piecewise constant function can approximate a nonlinear response with sufficient accuracy. It is not unusual that with one single input variable such as wind speed, people needs to use 20 bins to model the whole response. With 20 bins, the number of parameters cannot be fewer than 20, already producing a model that is unnecessarily complicated.

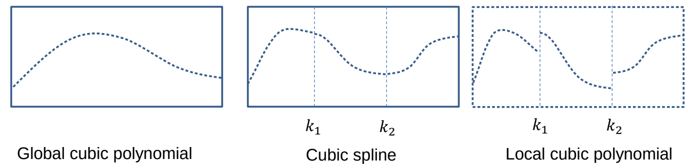


FIGURE 4. Global cubic polynomial, cubic spline, and local cubic polynomials.

As the number of inputs increases, the traditional spline methods would still run into the "curse of dimensionality" problem. A contemporary spline method, known as MARS standing for Multivariate Adaptive Regression Splines [9], was invented and it injects the scalability into the traditional spline-based approach. For an individual element in $\mathbf{x}$, say x_i, and at a given knot k, MARS uses two types of basis function: $[x_i - k]_+$ and $[k - x_i]_+$, where $[\cdot]_+ := \max(\cdot, 0)$. These two basis functions can be altogether denoted as $\pm[x_i - k]_+$. For a pair of two elements in $\mathbf{x}$, say x_i and x_j, their interaction can be modeled by the product of the individual basis functions, namely $[\pm(x_i - k)]_+[\pm(x_j - k)]_+$. The final MARS model approximates a function, say $f(\mathbf{x})$, by a linear combination of the basis functions, i.e.,

$$\hat{f}(\mathbf{x}) = \sum_{\ell=1}^{L} \beta_\ell B_\ell(\mathbf{x}), \tag{2}$$

(3)

where $B_\ell(\cdot), \ell = 1, \ldots, L$, are the basis functions taking the aforementioned forms, β_ℓ's are the coefficients of the basis functions, and L is the number of the basis functions. Fitting of the MARS model is to estimate the coefficients β_ℓ using a set of observed data pairs $\{\mathbf{x}, y\}$'s. The non-zero values in the coefficients β_ℓ can also inform practitioners about the important factors or factor interactions in affecting the response.

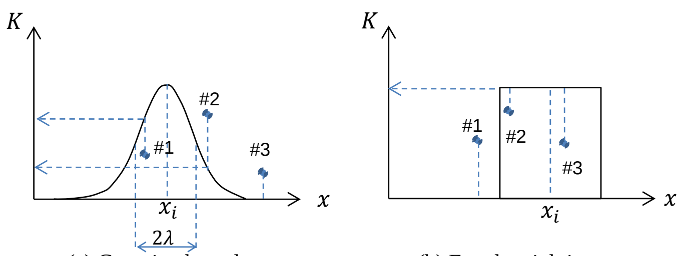
In order to make sure a MARS model effective, cares need to be exercised in deciding the number of knots used and their positions. As one can easily imagine, the choice of knots (both their number and positions) is equivalent to an adaptive binning operation. Understandably, the region where the response changes rapidly should have more knots with close proximity, whereas the region where the response changes slowly needs fewer knots with greater distance in between. The fewer the number of knots, the simpler the resulting model is, and the less pressure it places on the amount of training data required to yield a decent estimate. To this end, a Bayesian version of the MARS model was invented in [10, 11], and the Bayesian MARS model can decide the number and locations of knots automatically according to the specific dataset used. Recently, [2] uses a Bayesian MARS in their extreme load analysis of wind turbine bending moment data.

Kernel methods

The basic idea of a kernel method [12] is to make an estimation, $\hat{y}$, at x_i using observed data points close to x_i. This localization is achieved by employing a weighting function symmetric with respect to x_i, known as a kernel function and denoted by $K(x, x_i)$. A kernel function is supposed to be integrable to 1, so that the magnitude of $\hat{y}$ is consistent with that of original data y. The kernel function has a bandwidth parameter λ that controls how fast the function decays from its peak towards zero and determines the size of the local window. One of the popular kernel functions is the Gaussian kernel function, taking the form of

$$K_\lambda(x, x_i) = \frac{1}{\sqrt{2\pi\lambda^2}} \exp\left(-\frac{||x - x_i||^2}{2\lambda^2}\right). \quad (4)$$

Consider the example in the left panel of Figure 5 in which we see three data points. Points #1 and #2 have positive weights associated with them, while Point# 3 has a virtually zero weight. So the estimation of y at x_i is the weighted average of y values at points #1 and #2, respectively. Other data points that are farther away from x_i than #3 will not affect the estimation of y at x_i. In order to produce an estimate of y at a new location x_j, we simply move the kernel function to x_j and use the set of data points newly covered by the weighting function to make the new estimate.



(a) Gaussian kernel (b) Equal weighting

FIGURE 5. Gaussian kernel function versus step function used in binning.

Comparing the kernel model with the two ends of model spectrums, we can see their connections. The bandwidth parameter λ controls how locally a kernel model concentrates. Using a large enough λ, one could end up with using a single Gaussian density function to cover the whole input domain; this is the global version of the kernel model. On the other hand, using a small enough λ, one can let the kernel model localize as much as one wants. Comparing the kernel model with the binning method, one notices that the binning method can be considered as a special kernel model but uses a square pulse function as the weighting function, giving equal weights to all data points within the step function window, regardless how far away they are from x_i (points #2 and #3 in the right panel of Figure 5). Once a data point is outside the step function window (point # 1), its weight is zero. The final estimate at x_i is a simple average of all y's associated with the data points within the window. Of course, the binning method is not really a kernel model due to another important reason: in the kernel regression, the kernel function moves continuously along the x-axis, producing a continuous, smooth curve, while the step functions in the binning method are disjoint, so that the resulting function response from the binning method, if magnified enough, is discretized.

In the above narrative, when we use the "weighted average" of y values of the data points falling under the kernel function, the resulting outcome is the estimate of the mean of the conditional density $p(y|x)$ or simply $E(y|x)$. In practice, this mean function is useful enough, as it corresponds to the conventional power curve, should y be the power output of a turbine. The kernel methods, nonetheless, are capable of producing the estimate of the conditional density $p(y|x)$ also, through a formula like

$$\widehat{p}(y|x) = \sum_{i=1}^{N} w_i(x) K_{\lambda_y}(y - y_i), \quad (5)$$

where

$$w_i(x) = \frac{K_{\lambda_x}(||x - x_i||)}{\sum_{i=1}^{N} K_{\lambda_x}(||x - x_i|)}, \quad (6)$$

λ_y and λ_x are the bandwidth parameters associated with the response and the input variable, respectively, and N is the size of the data set used to estimate the conditional density.

Similar to the binning method and the plain version of spline methods, when encountering a multivariate analysis of which the number of inputs are more than one dimension, the kernel methods, if unmodified, will also run into the problem of "curse of dimensionality." We proposed in [3] an additive-multiplicative kernel (AMK) structure that strikes a sensible balance between scalability and the modeling ability in terms of capturing the factor interaction effects.

Suppose that one has q elements in the input vector $\mathbf{x}$, i.e., $\mathbf{x} = (x_1, \cdots, x_q)$. The AMK forms a series multiplicative trivariate kernel function taking an input vector of three dimensions, denoted by, for instance, $\tilde{\mathbf{x}}^{(1,2,3)} := (x_1, x_2, x_3)$, or generally, $\tilde{\mathbf{x}}^{(i,j,\ell)} := (x_i, x_j, x_\ell)$. One can use the multiplicative trivariate kernels to capture the interaction effect up to three-factor interactions. Then, all the trivariate kernels are pooled together in an additive structure:

$$\widehat{p}(y|\mathbf{x}) = \sum_{i=1}^{N} \frac{1}{Q} \left[w_i(\tilde{\mathbf{x}}^{(1,2,3)}) + w_i(\tilde{\mathbf{x}}^{(1,2,4)}) + \ldots + w_i(\tilde{\mathbf{x}}^{(i,j,\ell)}) \right] K_{\lambda_y}(y - y_i), \tag{7}$$

where Q is the number of the trivariate kernel functions in the additive model. Usually there is no need to enumerate all possible three-factor interactions, so that the value of Q can stay manageable. With this additive-multiplicative structure, AMK models can capture important nonlinear effects and their interactions, while staying scalable in practice.

WIND ENERGY RELEVANCE AND APPLICATIONS

This section presents two case studies: an extreme load analysis and a quantification of turbine upgrade using power curve.

CASE 1: Extreme load analysis

In this application, y is the maximum bending moments measured at the critical spots on a turbine structure (most likely blades). The definition of extreme load is the extreme quantile value in the distribution $p(y)$ corresponding to a turbines service time of T years, such as $P_T = Pr[y > l_T]$. International Electrotechnical Commission (IEC) [6] publishes P_T for a given T; for instance, when $T = 50$ years, $P_T = 4 \times 10^{-7}$. Once P_T is specified, l_T can be determined based on the distribution of y. Here l_T is the designed load endurance level, a design parameter used to select materials and the associated manufacturing process, so that as long as the actually experienced extreme load does not exceed l_T, the turbine's structural integrity will not be compromised. The specification of P_T is to ensure that the load threshold l_T may be exceeded only in a very small probability during a turbine's service life. The load threshold l_T is also called the extreme load level.

In the extreme load analysis, the distribution $p(y)$ is computed by integrating the conditional density $p(y|\mathbf{x})$ and weather profile distribution $p(\mathbf{x})$, as expressed in equation (1). The latter has been studied extensively by statisticians and meteorologists alike [13]. So the focus in [2] is to present a new model to handle $p(y|\mathbf{x})$.

As an extreme quantile value is sought after in the extreme load analysis, a generalized extreme value (GEV) distribution [14] is employed. A GEV distribution has three parameters: the location parameter μ, the scale parameter σ, and the shape parameter ξ; denoted by $GEV(\mu, \sigma, \xi)$. Depending on the choice of ξ, a GEV may manifest in a distribution more familiar to the practitioners: when $\xi > 0$, the GEV is a Fréchet distribution; when $\xi < 0$, the GEV is a Weibull distribution; when $\xi = 0$, the GEV is a Gumbel distribution.

The nonlinearity and heteroscedasticity of the wind-versus-load relationship demands a non-homogeneous modeling of the GEV distribution over the whole weather profile. The binning method is to partition the input domain into small bins, and then fit a constant parameter GEV distribution to each one of the bins. Suppose that we have the wind speed as the only input, and we partition the wind speed from the cut-in speed to the cut-out speed by a resolution of 1 m/s; this may result in about 20 bins. Then, we fit 20 individual GEV distributions, namely $GEV(\mu_1, \sigma_1, \xi_1), \ldots, GEV(\mu_{20}, \sigma_{20}, \xi_{20})$. Within each bin, the parameters (μ_i, σ_i, ξ_i) are constant, so that the GEV in that bin is homogeneous and stationary. Usually, people would let $\xi_1 = \xi_2 \ldots = \xi_{20} = \xi$ to reduce to some extent the flexibility of the model; otherwise the model may have too many degrees of freedom, leading to an underdetermined system with no unique solutions. The essence of the binning method is to handle a challenging non-homogeneous problem with a set of piecewise homogeneous GEV distributions.

What was proposed in [2] is to abandon the bins altogether, but instead develop a non-homogeneous GEV distribution over the entire input domain, supported by locally adaptive spline models. The locally adaptive spline models connect all the bins across the input region, so that the limited load and weather data are pooled together, allowing efficient and effective modeling and analysis.

Specifically, the modeling elements are as follows. First, a non-homogeneous GEV distribution is used, which means that its location parameter $\mu(\mathbf{x})$ and scale parameter $\sigma(\mathbf{x})$ are both a function of the input $\mathbf{x}$. Its shape parameter ξ can be a function of $\mathbf{x}$, too, but to follow the same treatment in the current practice, the shape parameter is kept as a constant over the whole input domain and to be estimated from the data. Consequently, the conditional distribution of $y|\mathbf{x}$ is expressed as $y|\mathbf{x} \sim GEV(\mu(\mathbf{x}); \sigma(\mathbf{x}); \xi)$; and $\sigma(\mathbf{x}) > 0$.

The two functions $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$ in the non-homogeneous GEV distribution are each modeled by a Bayesian MARS model, as explained in the previous section. The Bayesian MARS models are solved using a numerical sampling procedure called reversible jump Markov chain Monte Carlo (RJMCMC) [15]. The details of the model and the RJMCMC procedure can be found in [2].

Naturally, people ask the question – what difference does this MARS modeling technique make? First, it makes a drastic difference in terms of the complexity of the resulting models. Although the modeling procedure of Bayesian MARS is more involved than that of the binning method, the resulting MARS model in fact uses fewer parameters. In analyzing the wind-to-load data shown in the left panel of Figure 1, two wind-related covariates are used: the average wind speed and the standard deviation of wind speed (related to turbulence intensity). With two input variables, [2] tried a binning method that uses 60 bins, and each bin is fit with a stationary GEV distribution having two parameters (μ and σ). This translates to a total of 120 parameters for the resulting bin-based GEV model. By contrast, the MARS-supported GEV model uses about 20 parameters, nearly an order of magnitude fewer than that in the binning model. Second, the MARS-supported GEV model produces an estimate of the extreme load level l_T that is consistently lower than that by the binning method; see Figure 6 for an illustration. The estimation of the extreme load level is for $T = 20$ years. In [2], we conducted a simulation study in which l_T can be observed with sufficient amount of samples, so that the estimations from competing methods can be compared and assessed of their bias. It was confirmed that the binning method indeed has the tendency to overestimate the extreme load level, while the spline-supported GEV model produces estimations more in line of the observed l_T. These studies demonstrate that the overestimate by the binning method does not happen by chance. So [2] concluded with a statement like "*In the end, the spline method uses a sophisticated procedure to find a simpler model that is more capable.*"

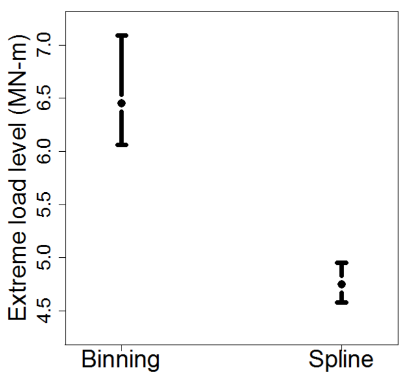


FIGURE 6. Comparison of the extreme load estimation using the bin-based GEV model and the spline-based GEV model. The middle point represents the mean of the extreme load estimate, while the two extreme points correspond to the 95% credible (or confidence) intervals. In [2], a simulation study in which l_T can be observed with sufficient amount of samples is conducted and confirms the observation that the binning method tends to overestimate the extreme load level.

CASE 2: Quantification of turbine upgrade

In this application, y is the the power output measured at individual turbines. This, together with the wind measurement in $\mathbf{x}$, leads to the power curves that are commonly used for power prediction [16, 17]. They can also be used for characterizing energy production efficiency of a wind turbine generator, because a change in the position and slope of a turbine's power curve, especially in the part between the cut-in wind speed and rated speed, indicates a change in energy production efficiency.

In this case study, we first estimate the power curve associated with a wind turbine. Technically, the objective is to estimate the conditional density, $p(y|\mathbf{x})$, or the conditional expectation, $E(y|\mathbf{x})$. Our analysis here, following the methodology presented in [3], is to build a so-called *endogenous power curve*, the curve decided primarily by a turbine's own aerodynamics, after the influence from the ambient environmental factors is controlled for. Towards that end, we use the AMK model to capture the multivariate dependencies and their interactions from an array of weather covariates including wind speed, wind direction, air density, humidity, turbulence intensity, above-hub wind shear and below-hub wind shear. Note that due to lack of instrumentation, not all the variables are available in every wind farm study conducted. If all included, the number of variables in $\mathbf{x}$ is seven, namely $q = 7$. With multiple dependencies, $E(y|\mathbf{x})$ is no longer a power curve but 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.

When using the AMK model framework and if we exhaust all the trivariate kernels, the total combinations would be $\mathbb{C}_7^3 = 35$. Having $Q = 35$ terms in an AMK is something solvable but may not be necessarily. The physical understanding of the wind power production tells us that wind speed and wind direction are two most important factors, whereas other factors may or may not interact with them. With this understanding, we decide to order the variables in $\mathbf{x}$ based on its likelihood of affecting the wind power production. This places wind speed as x_1, wind direction as x_2, air density as x_3, and so on. Then we keep the trivariate kernel terms that have x_1 and x_2. This yields a total of five terms, namely $\tilde{\mathbf{x}}^{(1)} = (x_1, x_2, x_3)$, $\tilde{\mathbf{x}}^{(2)} = (x_1, x_2, x_4)$, ..., $\tilde{\mathbf{x}}^{(5)} = (x_1, x_2, x_7)$. The number of terms in the AMK is then $Q = q - 2$, much smaller than the exhaustive list $\mathbb{C}_q^3$.

This endogenous power curve is then used to detect and quantify an upgrade undertaken on a wind turbine. Recognizing the fast deterioration of wind turbines as a result of working under harsh conditions and random cyclic stresses, wind operators from time to time perform certain retrofit to their turbine generators with the hope to restore or even boost its energy production efficiency. This action is called turbine upgrade. One of such retrofit is called the vortex generator (VG) installation [18, 19]. But VG installation is not inexpensive. Naturally, wind operators would like to know how much benefit can be expected from such an upgrade under actual operation conditions.

One wind operator asked us to conduct a blind test for detecting and quantifying turbine upgrades. The upgrade type is indeed vortex generator installation. Three turbines on the same wind farm were chosen. It was understood that some of the turbines may have undergone a VG type of upgrade. But the following information was withheld from us, namely that whether this VG upgrade actually took place, and if so, on which turbine(s) it was done. The specific date of a VG upgrade, if it indeed occurred, was made available to us. Fourteen months worth of turbine power data and weather measurement data were gathered on the wind turbines and were also provided to us. We are asked to find out whether there is an upgrade on any of the turbines and if the answer is positive, which one(s). This detection aspect can be verified with the knowledge of the wind operator who knew which turbine(s), if any, has been upgraded. If this answer is correct, then the next question is how much improvement this upgrade produces in terms of power production improvement.

We proceed to use both the binning method and the AMK method to fit the power curves associated with the turbines and compare the curves before and after the upgrade date. We conduct a statistical t-test [20] and compute the corresponding p value. According to the statistical convention, a p value smaller than 0.05 indicates a significant difference before and after the upgrade date, pointing to the occurrence of actual upgrade. A large p value suggests no upgrade. Table 1 presents our analysis outcome.

TABLE 1. Binning method and kernel method used for detecting a VG upgrade.

	Binning		Kernel	
	t statistic	p-value	t statistic	p-value
Turbine #1	-1.53	0.13	1.75	0.08
Turbine #2	-10.40	4.0×10^{-24}	1.27	0.20
Turbine #3	-6.22	7.4×10^{-10}	2.17	0.03

Based on our analysis using the kernel method, we informed the wind operator that there was a single turbine that underwent the VG upgrade and the upgraded turbine is #3. This was confirmed by the wind operator as the correct answer. The kernel-based method also suggests that after the upgrade, the power production efficiency is increased by 2.48%. This value is not easy to be validated experimentally, as it is difficult to control the environmental variables to be the same before and after an upgrade. Through a simulation study reported in [4], it seems that the kernel-based method was able to estimate the simulated improvement accurately.

Had we used the binning method, it would have flagged both Turbines #2 and #3 as the upgraded turbines. In fact, the confidence to say so about #2 is much stronger than that of #3, as the corresponding p-value is much smaller. We observed this type of erratic behavior of using the binning method previously in [4], and believe that such outcome is largely attributable to the still large amount of uncertainty unaccounted for while using the method.

CONCLUSION

This paper presents a discussion of conditional density estimation problem and its relevance to two types of wind energy applications. We discuss two categories of data analytic methods: the spline methods and the kernel methods. Both of them are generally classified as semi-parametric or nonparametric approaches, implying their flexible modeling capability and adaptivity to data. We recalled two wind energy-relevant analyses in which the two data analytic methods are used respectively. The studies using the actual wind turbine operational data (including the weather measurement data) appear to support the merit of adopting these contemporary data analytic methods in wind industry practices over the binning method, which has dominated the wind industry as the default data analytic tools for at least a decade.

ACKNOWLEDGMENT

Ding and Tang were partially supported by the grants from NSF (CMMI-1300560 and CMMI-1300236). Ding and Huang were partially supported by the grant from King Abdullah University of Science and Technology (KUS-CI-016-04).

REFERENCES

[1] DOE, 2014. Electric power monthly - data for July 2014 (release date: September 25, 2014). Tech. rep. Available at http://www.eia.gov/electricity/monthly/.

[2] Lee, G., Byon, E., Ntaimo, L., and Ding, Y., 2013. "Bayesian spline method for assessing extreme loads on wind turbines". *The Annals of Applied Statistics*, ***7***, pp. 2034–2061.

[3] Lee, G., Ding, Y., Genton, M. G., and Le, X., 2014. "Power curve estimation with multivariate environmental factors for inland and offshore wind farms". *Journal of the American Statistical Association*, ***accepted***.

[4] Lee, G., Ding, Y., Xie, L., and Genton, M. G., 2014. "Kernel plus method for quantifying wind turbine upgrades". *Wind Energy*, ***online published.***

[5] Agarwal, P., and Manuel, L., 2008. "Extreme loads for an offshore wind turbine using statistical extrapolation from limited field data". *Wind Energy*, ***11***, pp. 673–684.

[6] IEC, 2005. *Class of IEC international standards regarding wind turbines. IEC 61400-1: Design Requirements and IEC 61400-12: Wind turbine power performance testing*. International Electrotechnical Commission, Geneva, Switzerland.

[7] Hastie, T., and Tibshirani, R., 1990. *Generalized Additive Models*. Chapman and Hall, London, UK.
[8] Hastie, T., Tibshirani, R., and Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, New York.
[9] Friedman, J. H., 1991. "Multivariate adaptive regression splines". *The Annals of Statistics,* ***19***, pp. 1–67.
[10] Denison, D., Mallick, B., and Smith, A., 1998. "Bayesian MARS". *Statistics and Computing,* ***8***, pp. 337–346.
[11] Denison, D., Holmes, C., Mallick, B., and Smith, A., 2002. *Bayesian Methods for Nonlinear Classification and Regression*. Wiley, Chichester, UK.
[12] Loader, C., 1999. *Local Regression and Likelihood*. Springer, New York.
[13] Li, G., and Shi, J., 2010. "Application of Bayesian model averaging in modeling long-term wind speed distributions". *Renewable Energy,* ***35***, pp. 1192–1202.
[14] Smith, R., 1990. "Extreme value theory". In Handbook of Applicable Mathematics, Vol. 7, John Wiley, pp. 437–471.
[15] Green, P., 1995. "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination". *Biometrika,* ***82***, pp. 711–732.
[16] Monteiro, C., Bessa, R., Miranda, V., Botterud, A., Wang, J., and Conzelmann, G., 2009. Wind power forecasting: state-of-the-art 2009. Tech. Rep. ANL/DIS-10-1, Illinois, U.S. Available at http://www.osti.gov/energycitations/product.biblio.jsp?osti_id=968212.
[17] 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. Tech. rep., Roskilde, Denmark. Available at http://www.anemos-plus.eu/images/pubs/deliverables/aplus.deliverable_d1.2.stp_sota_v1.1.pdf.
[18] Timmer, W. A., and van Rooij, R. P. J. O. M., 2003. "Summary of the Delft University wind turbine dedicated airfoils". *Transactions of the ASME, Journal of Solar Energy Engineering,* ***125***, pp. 488–496.
[19] Øye, S., 1995. "The effect of vortex generators on the performance of the ELKRAFT 1000 kw turbine". In *Aerodynamics of Wind Turbines: 9th IEA Symposium*. Stockholm, Sweden, pp. 9–14. ISSN:0590-8809.
[20] Casella, G., and Berger, R. L., 2001. *Statistical Inference*. Cengage Learning, Independence, KY.