A Multi-Dimensional Power Curve Model

Dr. Yu Ding

Mike & Sugar Barnes Professor
Professor of Industrial & Systems Engineering
Professor of Electrical & Computer Engineering

Texas A&M University

September 29, 2016

What we did

- We developed a multi-dimensional power curve method that can account for the effects of multiple environmental factors (wind speed included).
- Our method is based on a machine learning technique (originated in the field of statistics), known as kernel regression or kernel density estimation.
- Because of the machine learning capability, we do not explicitly distinguish the standard versus non-standard wind conditions. Rather our model uses the data from both conditions and let the data guide the final model building.
- We found the new power curve model can significantly reduce the power prediction error as compared to the IEC standard method.
- Our method has an inherent uncertainty quantification capability.

Lee, Ding, Genton and Xie (2015) "Power curve estimation with multivariate environmental factors for inland and offshore wind farms," *Journal of the American Statistical Association*, Vol. 110, pp. 56-67.

What is kernel regression?

Kernel regression is one localized regression method, driven directly by data. The idea is to make a prediction at $\mathbf{x}_0$ using observations close to $\mathbf{x}_0$. This localization is achieved via the use of a weighting function, also known as the kernel function $k_\lambda(\mathbf{x}_0, \mathbf{x})$.

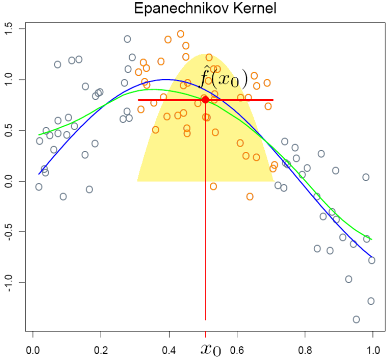


Figure: Credit: Hastie, Tibshirani, and Friedman (2009), The Elements of Statistical Learning, Springer.

- The blue curve is the underlying function unknown to a data analyst but to be recovered based on the noisy measurement data (the circles).
- The kernel function is visualized by the yellow shade; this particular kernel function uses a window width 0.4.
- The green curve is the estimated curve (or the recovered curve), which is computed by the weighted average of all data points falling into the kernel window (the shade area); these data points are shown in light brown color in the plot.

Our proposed method: AMK or Kernel PLUS

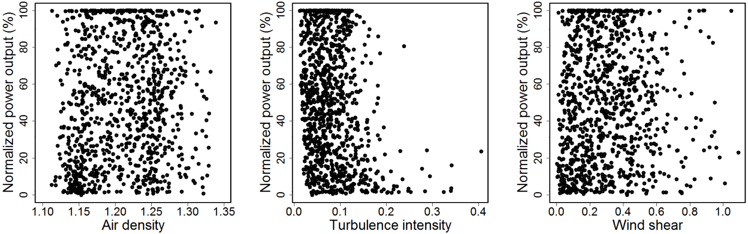
- Our multi-dimensional power curve method, referred to as **additive-multiplicative kernel (AMK)** in the research communities but nicknamed as **Kernel PLUS** for practitioners.
- Model inputs and output (all data arranged in 10-min):
  - **Output** y: the active power measured at a turbine;
  - **Input** $\boldsymbol{x}$: a vector, including the following elements:
    1. wind speed V;
    2. wind direction D;
    3. air density ρ;
    4. turbulence intensity S;
    5. humidity H;
    6. vertical wind shear W.
- Power prediction:

$$\hat{y}(\mathbf{x}) = \text{kernel}(V, D, \rho) + \text{kernel}(V, D, H) + \text{kernel}(V, D, S) + \text{kernel}(V, D, W) + \ldots$$

Other environmental variables can be included, if their field measurements are available.

Data used to train the model

- We test the method using actual wind turbine operational data, both offshore and inland. The typical duration of a dataset is one year, comprising about 30,000 - 40,000 data points.
- There are non-standard wind conditions in the training/test datasets.



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

- Offshore farm SEVEN variables: $\mathbf{x} = (V, D, \rho, S, H, W_a, W_b)^T$; Inland farm FIVE variables: $\mathbf{x} = (V, D, \rho, S, W_b)^T$.

W_a : above-hub wind shear; W_b : below-hub wind shear.

Model assessment criteria

- Out-of-sample testing: 80% data for training and 20% for testing.
- The training-test datasets are randomly split from the original dataset.
- Root mean square error (RMSE) for assessing prediction error:

$$\text{RMSE} = \sqrt{\frac{1}{N_{TS}} \sum_{i=1}^{N_{TS}} \left(\hat{y}(\mathbf{x}_i) - y_i\right)^2},$$

  where N_{TS} is the number of data pairs in the test data set.
- We refer to the IEC method as the **binning method**. Label the plain version without air density adjustment as **BIN** and the air density adjusted version as **BIN**$_a$.

Comparison

- (V, D, ρ) means that the Kernel PLUS has only one term (V, D, ρ): kernel(V, D, ρ), whereas (V, D, ρ, S) means that the model has two terms: kernel(V, D, ρ)+kernel(V, D, S). Same for (V, D, ρ, H).
- The percentage in the parenthesis is the improvement the Kernel PLUS made over the IEC binning method BIN_a (air density adjusted version).

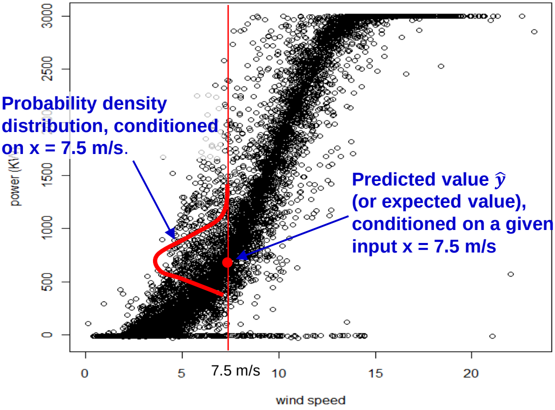
Wind Farm	Turbine	BIN_a	Additive multiplicative kernel		
			(V, D)	(V, D, ρ)	(V, D, ρ, S) or (V, D, ρ, H)
Inland	WT1	218.7	148.2 (32.3%)	124.7 (43.0%)	124.2 (43.2%)
	WT2	194.3	152.2 (24.8%)	130.4 (32.9%)	128.0 (34.1%)
	WT3	209.2	149.8 (28.4%)	122.2 (41.6%)	118.9 (43.2%)
	WT4	267.9	199.0 (25.7%)	175.3 (34.6%)	171.3 (36.1%)
Offshore	WT5	332.1	279.9 (15.8%)	252.6 (23.9%)	248.2 (25.6%)
	WT6	376.0	293.7 (21.9%)	261.6 (30.4%)	256.7 (31.7%)

An Observation

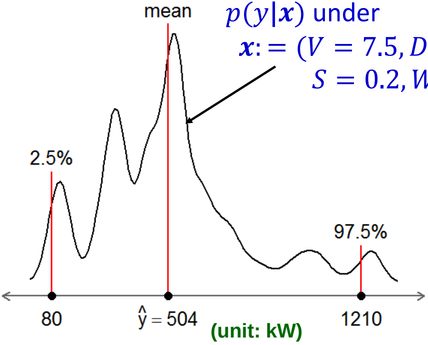
- Using the four inland turbine datasets, the following table compares the improvements made by including the factor of air density in the IEC procedure as well as in the Kernel PLUS method.
- The percentage in the parenthesis is the reduction compared to the respective baseline.
- Including air density makes a much greater improvement in Kernel PLUS, suggesting that not only is incorporating a relevant factor important, but the specific way of incorporating that factor is equally important.

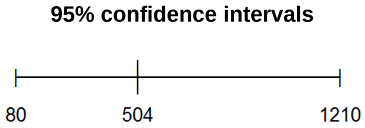
Wind Farm	Turbine	IEC Binning		Additive multiplicative kernel	
		Plain BIN	BIN_a	(V, D)	(V, D, ρ)
Inland	WT1	224.4	218.7 (2.8%)	148.2	124.7 (15.9%)
	WT2	202.8	194.3 (4.3%)	152.2	130.4 (14.3%)
	WT3	217.3	209.2 (3.8%)	149.8	122.2 (18.4%)
	WT4	280.5	267.9 (4.5%)	199.0	175.3 (11.9%)

- To quantify the uncertainty in predicting power output y, we need to model the conditional density function, $p(y|\mathbf{x})$, which is the probability density of y, conditioned on a given set of environmental conditions in $\mathbf{x}$.
- In fact, $\hat{y}(\mathbf{x})$ so far obtained is the **conditional expectation** $E(y|\mathbf{x})$, or the mean function of $p(y|\mathbf{x})$.

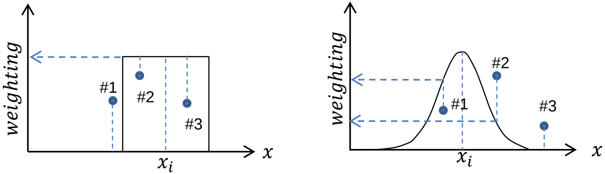


- Extending the Kernel model from estimating $\hat{y}(\mathbf{x})$ to estimating $p(y|\mathbf{x})$ is straightforward. Once $p(y|\mathbf{x})$ is estimated from the data, a confidence interval can be easily added to any predicted power output.





- IEC binning can be considered as a **special case** of Kernel PLUS:
  - Kernel PLUS: multi-dimensional, use a smooth function as its window;
  - Binning: one-dimensional, use a step function as its window.



(a) Binning: equal weighting (b) Kernel: different weighting

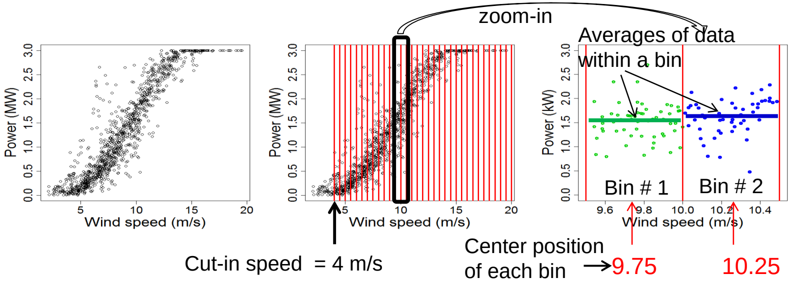
- **Why may it matter**?
  - Kernel PLUS significantly reduces the power prediction error over IEC binning, ranging from **27%** to **43%** on six turbine datasets.
  - Kernel PLUS can be used to create a 2-D Power Deviation Matrix, or a 3-D Power Deviation Matrix, or other high-D Power Deviation Matrices.

Part II

Appendix: IEC Binning Versus the Kernel Method

IEC standard method for power curve estimation

- In the current practice, the power curve is a one-dimensional curve, taking wind speed as input and power production as output.
- IEC method, and we call it the **binning method**, for estimating the power curve entails two steps:
  - Partition the range of wind speed into disjoint bins, using a bin width of 0.5 m/s;
  - Calculate the power response for a bin by averaging the power data falling in that bin.



Merit of binning method

Binning method provides a simple procedure to handle a difficult problem.

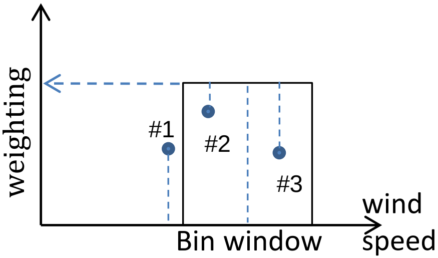
- The power curve is nonlinear in nature and does not have a fixed functional form. Even though people understand:

$$\text{Power} = \frac{1}{2}\rho \cdot A \cdot C_p \cdot V^3,$$

  this does not make the wind power output a cubic function of wind speed, because the power coefficient C_p is in turn a function of wind speed.

- Binning method is easy to understand, easy to implement, adaptive, and relies on few assumptions. In academia, we categorize methods like the binning method a "**nonparametric method**."

- The binning method uses a step function as its bin window. In the following plot, the data points (for example, #2, #3) in the bin window are given a weight of 1 and the data points outside the window (for example, #1) are given a weight of 0. Then the weighted average of all data points produces the power estimate for this particular bin.
- As the bin window is moving along the wind speed axis, the power curve is produced but not guaranteed to be smooth. Certain smoothing post-processing may be required.



Limitations of the binning method (continued)

- A **much greater challenge** for the binning method is its shortcoming in handling multiple input variables, leading to the problem known as "curse of dimensionality".
    - On the one hand, in addition to wind speed, other environmental variables also affect wind power production. So there is indeed a need to incorporate multiple input variables and extend the current one-dimensional power curve model to a multi-dimensional model.
    - On the other hand, it is not easy to do so within the current IEC framework, namely using the binning approach.
    - To see this, consider the following. Suppose that we now try to bin one year worth of 10-min data (about 52,000 data points) against five input variables, and each of them uses 20 bins. Then, we end up with more than **three million bins**. On average, every 60 bins share one data point.

What is kernel regression?

Kernel regression is one localized regression method, driven directly by data. The idea is to make a prediction at $\mathbf{x}_0$ using observations close to $\mathbf{x}_0$. This localization is achieved via the use of a weighting function, also known as the kernel function $k_\lambda(\mathbf{x}_0, \mathbf{x})$.

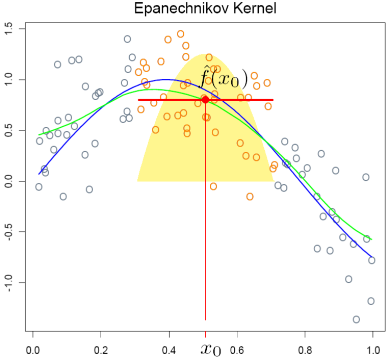


Figure: Credit: Hastie, Tibshirani, and Friedman (2009), The Elements of Statistical Learning, Springer.

- The blue curve is the underlying function unknown to a data analyst but to be recovered based on the noisy measurement data (the circles).
- The kernel function is visualized by the yellow shade; this particular kernel function uses a window width 0.4.
- The green curve is the estimated curve (or the recovered curve), which is computed by the weighted average of all data points falling into the kernel window (the shade area); these data points are shown in light brown color in the plot.

One-dimensional kernel

- Suppose we use a Gaussian kernel function

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

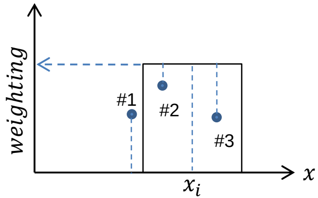
- The prediction formulation (known as Nadaraya-Watson kernel estimator):

$$\hat{y}(x) = \sum_{i=1}^{N} w_i(x) \cdot y_i, \text{ where } \quad w_i(x) = \frac{K_\lambda(x, x_i)}{\sum_{i=1}^{N} K_\lambda(x, x_i)}.$$

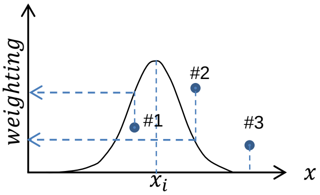
  - $\{x_i, y_i\}, i = 1, ..., N$ are N pairs of input-output observations.
  - $\hat{y}(x)$: predicted value at input x;
  - λ: bandwidth parameter = half window width;
  - $|| \cdot ||$: Euclidean distance.

Compare one-dimensional kernel with binning

- Recall that the binning method can also be considered as a weighted averaging method of data points falling into a window; see Slide # 15. The difference between binning and kernel is that they use different windows.
- In binning, the window is a step function. It gives all data points within the window the same weight and all outside points zero weight.
- In kernel, the window is a smooth function of a decaying shape, which decreases from its peak gradually to zero. The within-window data points are given different weights, depending at which location on the kernel function it is projected onto.



(a) Binning: equal weighting

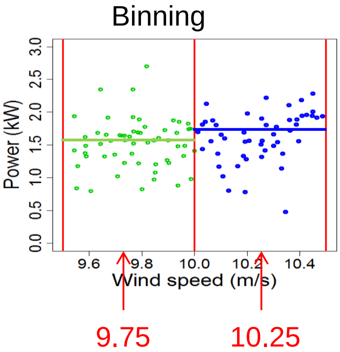


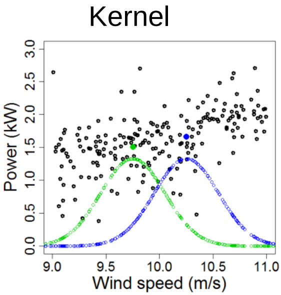
(b) Kernel: different weighting

Compare one-dimensional kernel with binning (continued)

Using a one-dimensional simple kernel estimate with wind speed as the only input, it produces a power estimate similar to that obtained via the binning method. In the example below,

- Binning method: For the two bins, estimates are 1.53 and 1.63, respectively.
- Simple kernel: The respective estimates are 1.51 and 1.66.





Multivariate kernel

- Multiple input variables in vector $\mathbf{x}$: $\mathbf{x} = (x_1, x_2, ..., x_q)$.
- A multivariate kernel function is a multiplication of univariate kernel functions, each of which bears the form on Slide # 18:

$$K_{\boldsymbol{\lambda}}(\mathbf{x}, \mathbf{x}_i) = K_{\lambda_1}(x_1, x_{i,1}) \cdot K_{\lambda_2}(x_2, x_{i,2}) \cdots K_{\lambda_q}(x_q, x_{i,q}),$$

  where $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_q)$ is the set of bandwidths associated with each covariate, and $x_{i,j}$ is the jth element of $\mathbf{x}_i$.
- A multivariate kernel regression is almost the same as the one-dimensional kernel regression, except that a multivariate kernel function needs to be used.
- But a multivariate kernel regression may suffer from the curse of dimensionality as well, if all univariate kernel functions are multiplied together to produce the multivariate kernel function.

Kernel PLUS handles multi-dimensional inputs

- Physical law provides us some clue.

Physics behind:

$$\text{Power} = \frac{1}{2}\rho \cdot A \cdot C_p \cdot V^3$$

- At least three important factors affect wind power generation;
- Functional relationship nonlinear with function form unknown;
- Interactions exist among the factors.

- This insights motivate a hybrid structure that we name it an **additive multivariate kernel (AMK) model**:

$$\hat{y}(\mathbf{x}) \longleftarrow \text{kernel}(V, D, \rho) + \text{kernel}(V, D, H) + \text{kernel}(V, D, S) + \ldots$$

  - A three-component multivariate kernel captures the critical factor interactions (with wind speed and wind direction);
  - An additive structure ensures the scalability of the final model.

Estimating the conditional probability density

- Extending the Kernel model from estimating $\hat{y}(\mathbf{x})$ to $p(y|\mathbf{x})$ is straightforward, because the general kernel method was initially developed for the purpose of estimating a probability density function.
- Specifically, the density estimation is fulfilled by replacing observations y_i with a kernel function created for y, $K_{\lambda_y}(y - y_i)$, namely:

$$\text{Conditional density estimator} \quad : \quad \hat{p}(y|\mathbf{x}) = \sum_{i=1}^{N} w_i(\mathbf{x}) \cdot K_{\lambda_y}(y - y_i),$$

$$\text{Conditional expectation estimator} \quad : \quad \hat{y}(\mathbf{x}) = \sum_{i=1}^{N} w_i(\mathbf{x}) \cdot y_i,$$

- Once $p(y|\mathbf{x})$ is estimated from the data, a confidence interval can be easily added to any predicted power output; see the plot on Slide # 10.