Boosting Technique for Power Curve Estimation

Jay Shah, Bhaskar Botcha
Department of Industrial and Systems Engineering
Texas A&M University
College Station, TX-77840

Executive Summary

POWER curve is widely used in the wind power industry to estimate the output power generated given a a set of operating conditions. It is also widely used for analyzing the health of a turbine and the efficiency of a wind-farm . The widely used convention in the industry, the EIC binning method, considers the wind-speed as the predictor and estimates the power generated. The industry uses this as a 'gold standard', but many works have shown that wind speed alone is not sufficient for accurate estimations of wind power generated. A plethora of environmental factors like wind direction, air density, temperature also play a major role in the power generation of a turbine. Also, the power readings from a wind-farm show that the variability of wind power for a given speed and operating conditions is very high, suggesting that there is room for improvement in predicting the power curve. The present implementation uses a 'boosting' technique to effectively capture the intricacies in the power curve by using incremental improvements of a tree-based model.

In the present project, working as a group, we learnt a numerous things which were crucial to step up our knowledge base in machine learning. The course gave us important tools for analysis and prediction of data including basic methods like regression and classification to advanced methods like Gaussian process regression, tree based models, etc. Using our understanding of methods in class, we employed numerous (almost all) methods to the present data and understood how each type of learning methods behave with the data. Also, we took home the modeling approaches required to model real-life data, which depends on a numerous factors than just a few obvious variables considered for studies. We learnt to use ML techniques like feature engineering, variable selection, parameter tuning and modeling techniques like grid-search to enhance our predictive models. We also wanted to explore the field of deep-learning, by looking at it's wide deployment in the field of ML. Although deep-learning (a neural net framework) was not our go-to model used finally due to low prediction accuracies than the boosting model we developed, we got an understanding on how black-box model works. After exploration of various models (which will be shown in the comparative analysis section in the sequel), we chose Boosting as our final model. Although, boosting is also viewed as a black-box technique by many, we explored the effects of variables within a boosting model by using partial dependency plots and supporting our arguments on why the boosting model is to be used.

Index Terms

Power curve, Boosting technique, machine learning, renewable energy, power forecasting, feature engineering

I. Introduction

In the wake of growing demand for clean energy sources and stringent environmental regulations, demand and research on renewable energy sources in the rise. Among the total renewable energy sources, wind energy accounts to 4.7% of the energy generated in the United States and 1.9% of the energy consumption of the world. Few reports show that by 2050, wround 25-30 % of global power could come from harnessing the energy from wind [1]. To drive this need by building efficient and reliable farms, extensive work is being done on estimating the power generation and turbine efficiency, given a set of environmental conditions. Wind speed is known as the major contributor to prediction of wind power generated [2]. The functional form describing the relationship between wind speed and power is called a 'Power Curve'. A nominal power curve provided by a manufacturer is shown in Fig.1a.The cut-in speed is the speed at which the turbine starts power generation, the rated speed is the speed at which the turbine delivers maximum power(also called 'rated power'). The cut-out speed is the speed at which the turbine's power generation will be halted due to safety issues. The actual power curve from data collected in a wind farm is shown

TABLE I: Features from data provided

Feature	Range
Wind speed(V)	3.02m/s - 20.336m/s
Wind speed standard deviation(S)	0m/s - 3.175m/s
Wind Direction(D)	0.039°- 359.957°
Environmental temperature(T)	-3.201°C - 36.992°C
Power(P)	0.09 - 1

in Fig.1b. The power curve obtained from actual sensor data shows significant variation in the operating region, which is not captured by a simple power curve provided by the manufacturer. The variability suggests the effect of few other variables on the power generated.

The wind power industry uses a 'standard' called the IEC(International Electrotechnical Commission) binning method. The methods uses equally spaced bins to calculate an estimate of the average power within the bin. This method averages out all the variability within a bin and with sufficient bin size, almost approximates to the nominal power curve provided by the manufacturer which ignores effects of other environmental variables. Techniques like curve-fitting using sigmoid functions [3] and Gaussian CDFs [4] were used which are known to approximate the shape of the power curve with appropriate shape parameters, but still can't provide an explanation to the variability of observations.

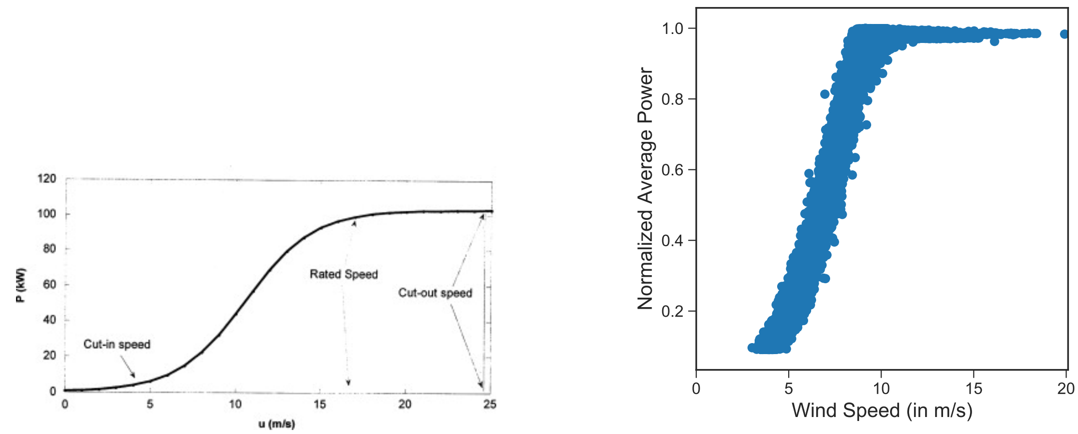
Sensors placed on current day masts are capable of capturing information like wind direction, temperature, air pressure and humidity. A few works point towards using wind direction along with wind speed as the explanatory variables for predicting wind power [5], [6]. Power curve estimations using multi-variate dependencies also have been noted in literature [7], [8], [9]. The works focus on how the models perform with multiple variables and their relative performance over standard traditional methods. For example, [7] used an additive multivariate kernel method for power curve estimation. [8] provided two data analytical methods, a Multivariate Adaptive Regression Spline(MARS) and an Additive Multiplicative Kernel (AMK) methods to predict wind power using multi-sensor data fusion approaches. Specific works using physical phenomenon like 'wake effects' were investigated in [10] and were futher modeled using spline methods. Recent works also implement techniques like Gaussian processes and Artificial Neural Networks (ANN) [11], [12]. A comprehensive review on different power curve modeling techniques can also be found in [13].

In our present implementation, we use a machine-learning technique called 'Boosting', which is a gradient boosting method(gradient because it uses residuals from one stage to build the subsequent model) based on multiple tree-based models. The usage of boosting for power curve estimation was not found in the present literature. Also, the method which was considered a black-box was analyzed for effect of the variables using tools like Partial Dependency Plots (PDPs) [14] and Individual Conditional Expectation (ICE) plots [15] which help visualize the relationships between the response and the feature-set. The independent variables used in our study are wind speed, wind direction, temperature, month and meridian. From the set of independent variables, a set of dependent variables, viz., air density and turbulence were extracted. Choice of features present and features extracted is explained in SectionII. The remainder of the paper is organized as follows: Section II describes the data and explains the dependencies between the variables and feature engineering depending on the physical phenomenon observed related to wind power prediction. Section III explains the boosting technique used for model development. Section IV discusses the results from the boosting model. Section IV-B provides a comparative analysis of different learning techniques employed in the present work and why boosting was used as the final model. The work is concluded in Section V.

II. Description and Visualization of data

A. Features and response variable from training data-set

All features provided for training the model with their respective range and the response variable 'Power' are shown in Table I.



(a) (b)

Fig. 1: (a)Power curve provided by turbine manufacturer. (b)Power curve from actual sensors located on a wind mast

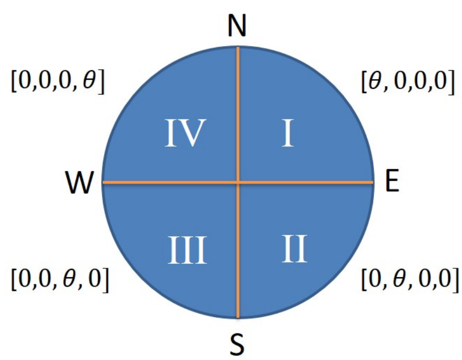


Fig. 2: Direction encoding

B. Feature engineering

As with many learning models, a few features may need to be removed or added based on the application. In the present implementation, a few features were added to the training set which we thought would increase the prediction accuracy of the model. First, a binary(0-1) response for the time of the day was created and is called 'meridian'(M). The choice of this variable is based on the simple reasoning that the temperature during the day are significantly higher than that during the night. This affects the air density and thus the wind speed. Also, a variable describing the month (Mo) was created. The reasoning follows along the same lines as the temperature change is significant with the month. The variation of temperature with month is shown in Fig.5.

Few dependent variables were also added to the feature set, viz., air density(ρ) and turbulence intensity (I), defined as follows,

- **Turbulence instensity,** I is calculated as the standard deviation of the wind speeds in a 10 min duration divide it with the average velocity within the time-frame. $I = \hat{\sigma_v}/V$
- **Air density,** ρ is calculated as $\rho = P/(RT)$ where P is the air pressure, R is the gas constant $R = 287Jkg^{-1}K^{-1}$ and T is the temperature in Kelvin.
- **Direction vector,** D_v, is a vector with four elements describing the direction of wind. (See Fig.2). The encoding makes sure that the direction of the wind is captured effectively. This was also verified with cross-validated results of the model. For subsequent analysis, the direction D is dropped off and the direction vector D_v is used.

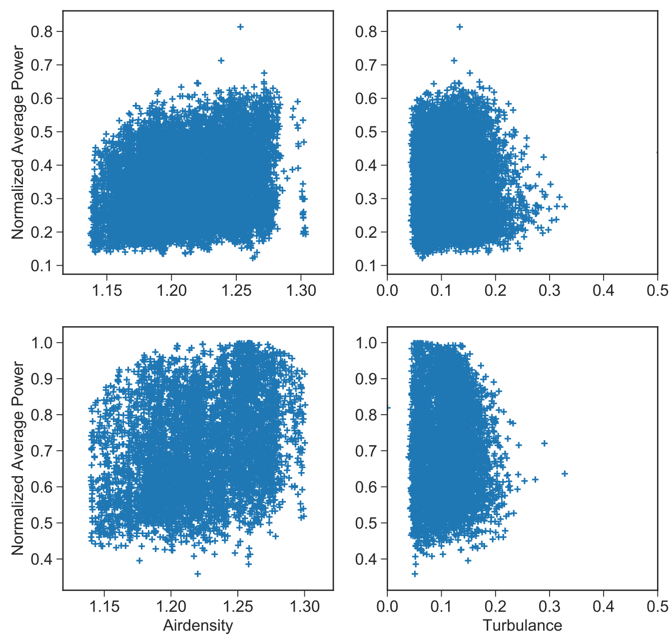


Fig. 3: Variation of power with air density and turbulence

The relationship between air-density

Finally, the power output can be expresses as a function of all the variables given and the extracted features as,

$$P_{avg} = f(V, S, D_v, T, M, Mo, I, \rho)$$

C. Trends

Analyzing trends in data and visually finding correlations is the first step in any modeling procedure. For the given data, Fig.4 shows the relationship between the explanatory variables. Also, Fig.5 shows the variation of temperature with month, which justifies the addition of month as a feature to the existing training data.

For a similar dataset, [7] has shown that the time-dependency of the data is negligible using auto-correlation and partial auto-correlation plots using a 1 hour lag. The reason for checking the trends in time-series is because, if the future predictions depends on the past observation, it has to be accounted for to get better predictions of the future. In the present model, the time dependency is omitted for further analysis.

III. Boosting model for power curve estimation

In this section, we introduce the Boosting technique and present the final model deployed for power curve prediction. Gradient boosting is a class of supervised learning methods which uses tree ensembles. The tree ensemble models are a set of Classification and Regression Tress (CART).

Ensemble methods, particularly trees, offer an attractive means to capture the underlying nonlinear relationships without imposing biased structures, as in conventional regression methods. They can also re-use the sparse experimental data to capture the empirical relationships connecting the features $X^{(i)}$ extracted based on the foregoing analytical model as well as experimental observations to predict the output(here, power average). Given a feature set $\psi_0 = (X_i, P_{avg}, i = 1, 2, ..N$ from experiments, the boosting model generates a tree model with a specified depth d. The residuals after fitting the first model, are considered a new data-set (ψ_1 after the first iteration) and a CART with the same depth is created. The process is repeated for Γ number of trees and the prediction is the sum of all predictions over Γ number of trees. The error-function used is a 'squared-error loss' function. The combination of

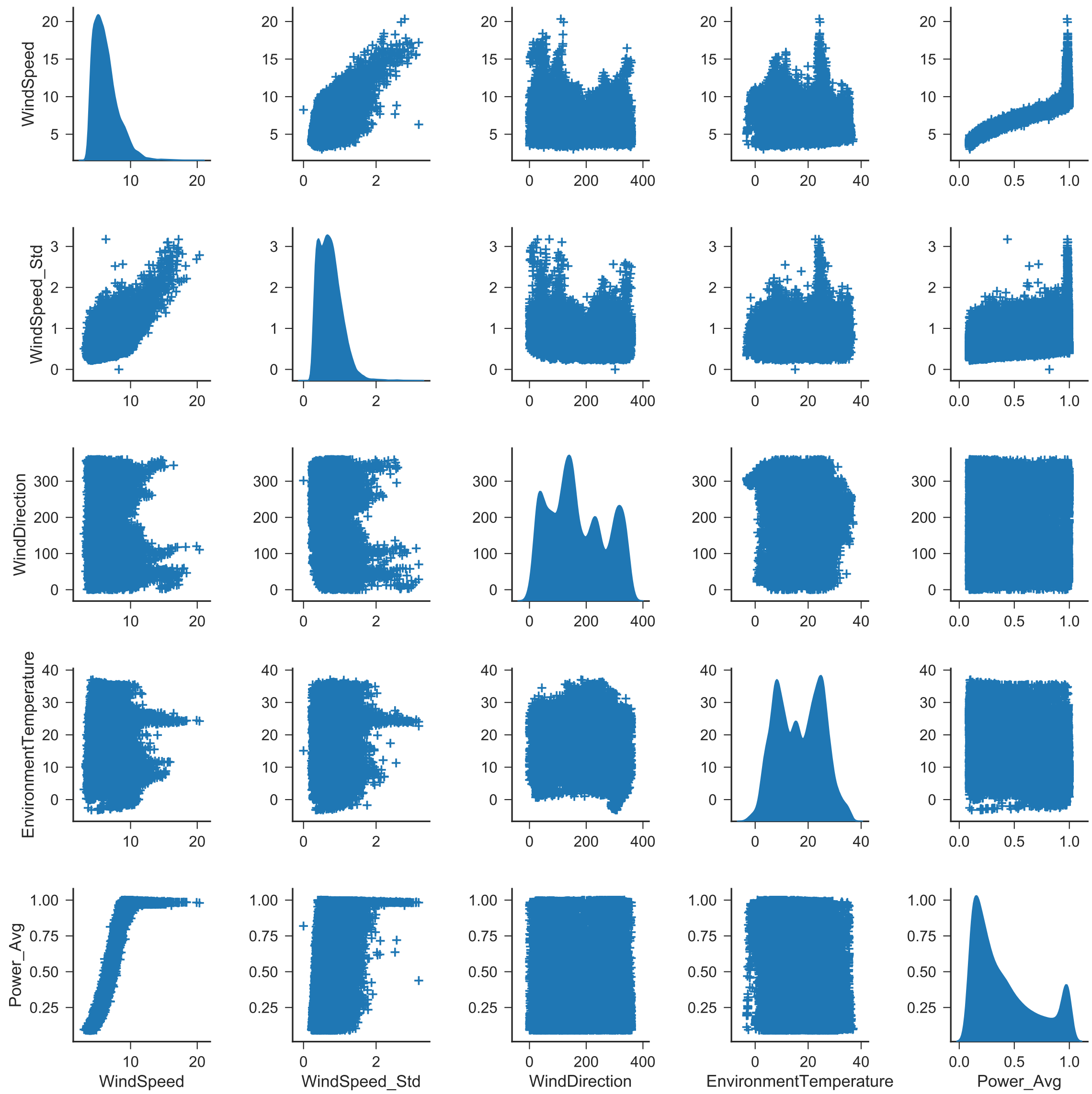


Fig. 4: Pair plots of variables used for training

aggregation of predictions from multiple (tree) models together with the use of different sets of input features to model each tree helps reduce variance and avoid over-fitting. A detailed explanation of the algorithm is shown in Fig.6. For the present implementation, a package called 'XGBoost' was used which is a fast and effective way to implement the boosting algorithm. The values of Γ, the number of trees and d, the depth of trees is obtained using 'grid search' technique, which covers the user-given range of parameters Γ and d and run multiple models until the collection of parameters with the least RMSE value is attained. The CV results for the number of trees and depth of trees are shown in Fig.7a and Fig.7a respectively.

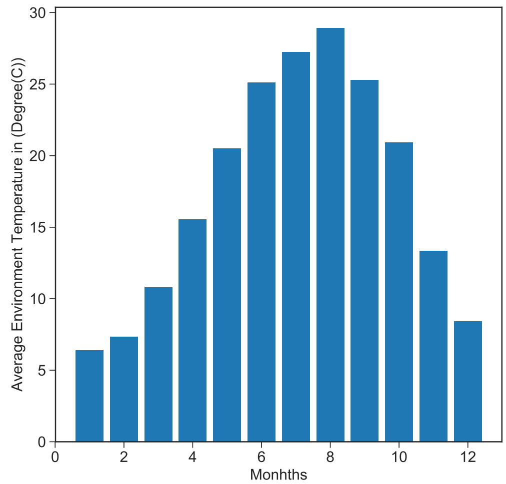


Fig. 5: Average temperature vs month

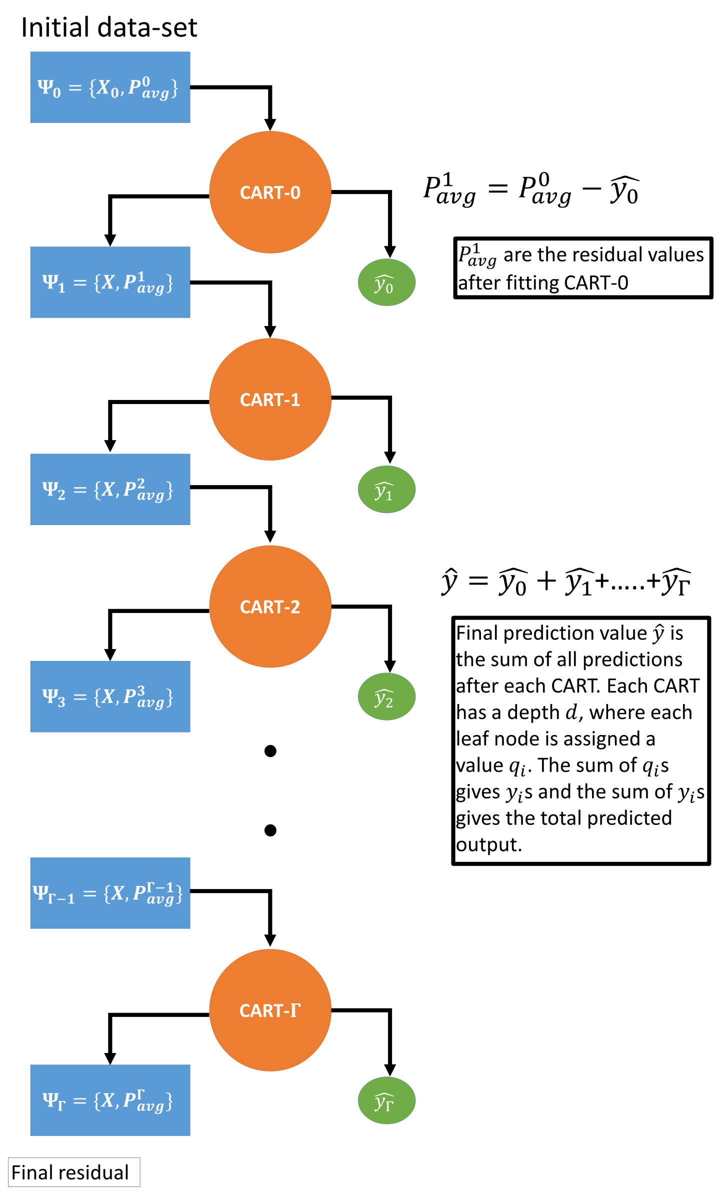
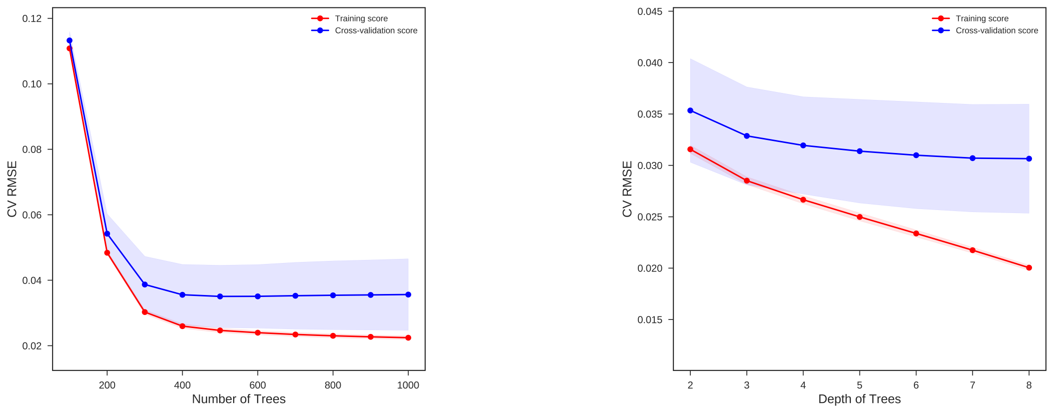
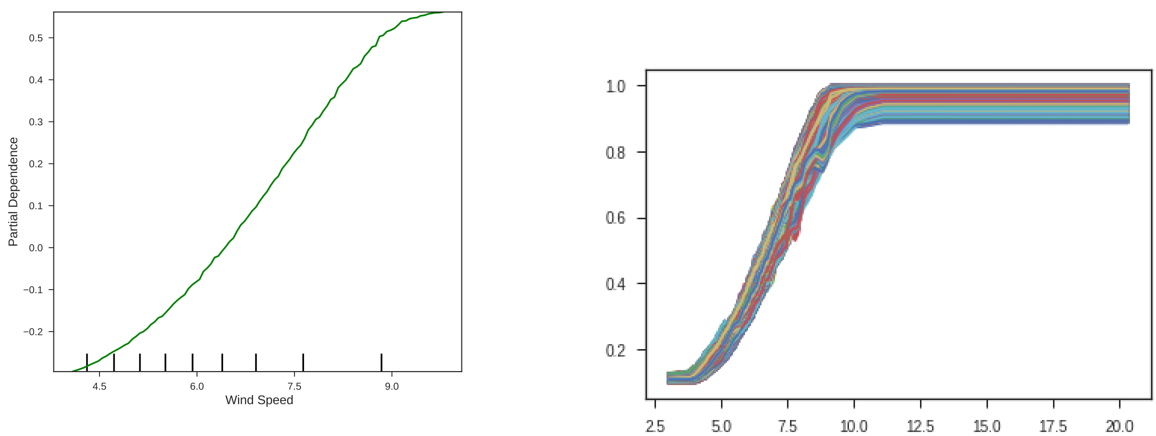


Fig. 6: Flowchart explaining the working of a Boosting method



(a) (b)

Fig. 7: (a) CV curve for number of trees (b) CV curve for maximum depth of each tree



(a) (b)

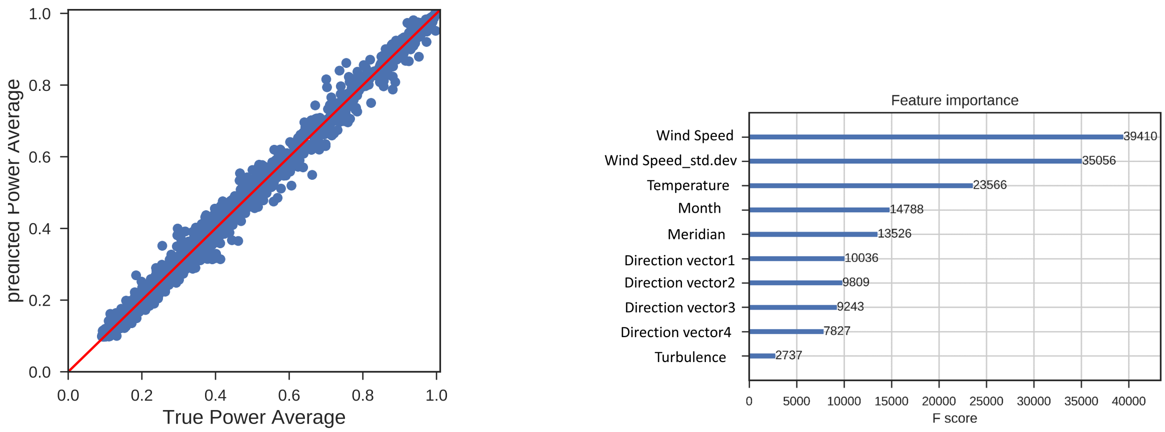
Fig. 8: (a) PDP plot for wind speed (b) ICE plot for wind speed

To interpret the model better, techniques like partial dependency plot are used. PDPs plot the change in the avergae predicted value as specific feature(s) vary over their marginal distribution [14]. PDP plots for wind speed is shown in Fig.8a. In addition to PDPs, ICE(Individual Conditional Expectation) plots [15] are generated. ICE plots N (the number of observations) estimated conditional expectation curves, where each reflects the predicted response as a function of a co-variate, conditional on an observation. The ICE gives insights into several invariants of the conditional relationship, while the PDP is just an average of all the expectations [15]. The PDP and ICE plots are shown in Fig.8a and Fig.8b respectively.

IV. Results

A. Results from the boosting model

The prediction results of our final model are shown in the Fig.9a. The results show that the predicted values follow very closely(in contrast to those in Fig.10) with the observed values from wind farm data. The parameters of the model are shown in Table II. The values are obtained through a grid-search which uses combination of parameter values over the parameter space and using a CV-result on all models to arrive at the model with the lowest RMSE. The relative importance of features is shown in Fig.9b.



(a))

(b)

Fig. 9: (a) Results from boosting model - Predicted vs observed values (b) Importance of features

TABLE II: Parameters of the boosting model

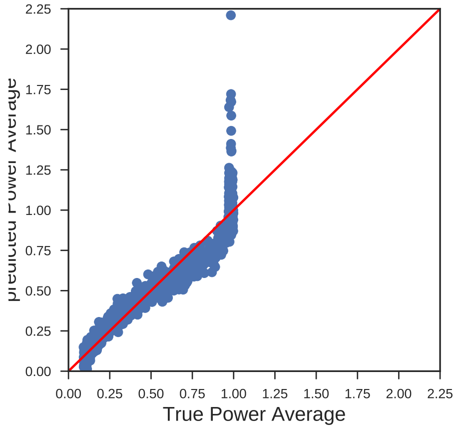
Parameter	**Value**
Maximum Tree depth d	8
Number of trees Γ	1000
Learning rate	0.01
Loss function L	RMSE

B. Comparative analysis

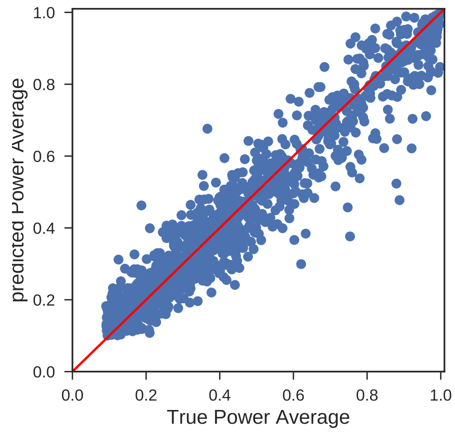
To study the relative merits and de-merits in modeling the present data, various models were built using different learning methods. The methods and their results are provided in this section and a final comparison of all methods is shown towards the end of this section. The following methods were used in the current study:

- Multiple linear regression
- K-Nearest Neighbors regression
- Support Vector regression
- Trees
- Random Forests
- Neural networks

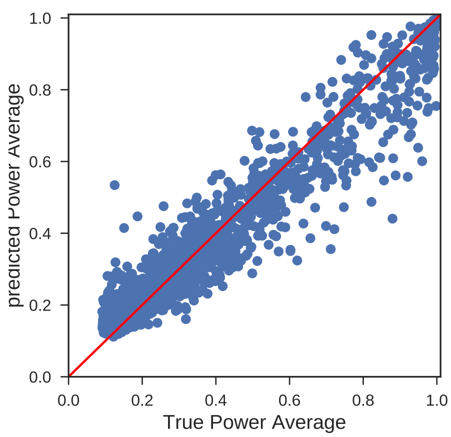
It can be seen from Fig.10 that the linear regression model has few predictions greater than 1, which can never be the case (Can't exceed rated power). Linear regression models generally fail when extrapolating outside the feature space. In the current study, we used cross-validated RMSE as a decision factor(best RMSE for a given model developed) to select the best model and Boosting was the model with minimum RMSE of 0.03. The comparison of different models can be seen in Fig.11.



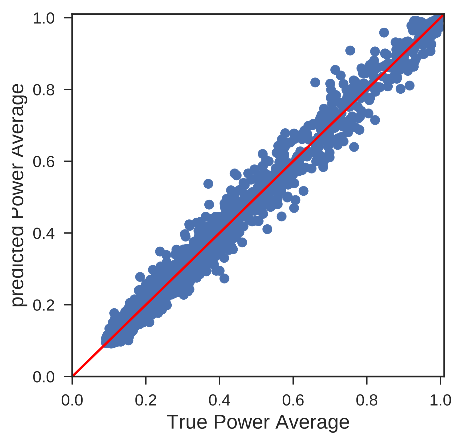
(a) Multiple Linear Regression



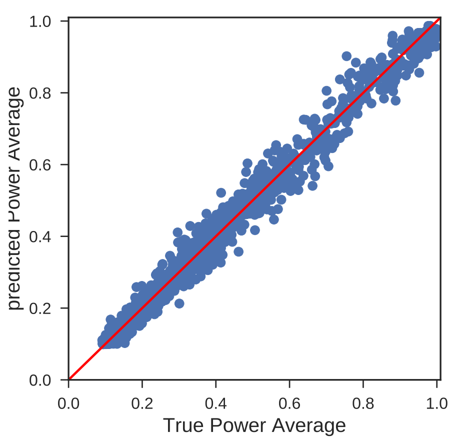
(b) K-nearest neighbors



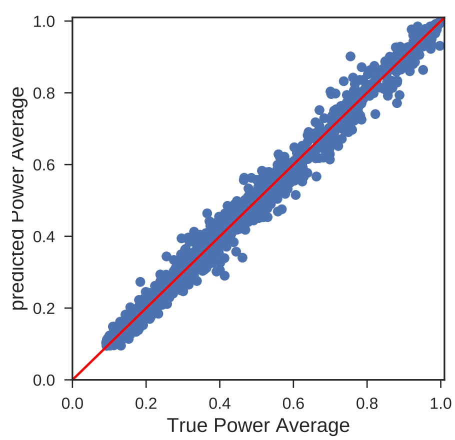
(c) Support Vector Machines



(d) Trees



(e) Artificial Neural Network



(f) Random forests

Fig. 10: Comparison of different models

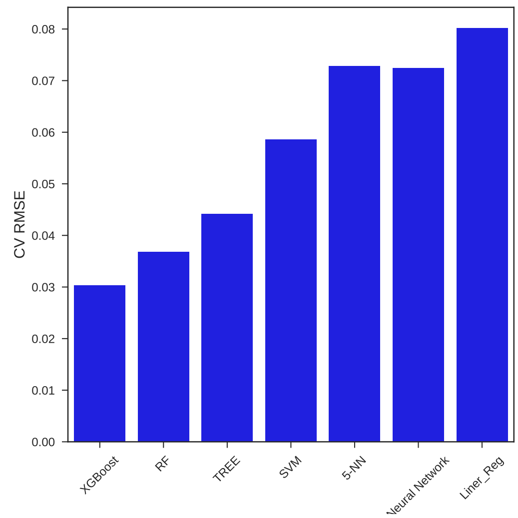


Fig. 11: Comparison of different learning methods

V. CONCLUSIONS

In the current effort to model wind turbine power curve, we presented a Boosting model as an effective model in contrast to many other learning methods discussed in SectionIV-B. Not viewing boosting as a complete black-box model, we have shown the how PDP and ICE can be used to find the effect of variables on the final prediction of power. Also, ensemble methods like boosting help capture non-linear relationships in data without imposing any structural form, thereby reducing the bias. The final results with cross-validated RMSE of 0.0303 show that the predictions are far better (1.5%) compared to the international standard in wind power curve prediction, the IEC binning method which is limited by the compartmentalization of data, using the data in the bin itself, but not the effect of other bins. Further, binning is a good method for 1-D data, but for data with multiple predictors and inter-dependencies, we have shown that ensemble methods perform better.

REFERENCES

[1] 2014. [Online]. Available: https://www.ecowatch.com/wind-energy-could-generate-nearly-20-percent-of-worlds-electricity-by--1881962962.html

[2] C. Monteiro, R. Bessa, V. Miranda, A. Botterud, J. Wang, G. Conzelmann *et al.*, "Wind power forecasting: state-of-the-art 2009." Argonne National Laboratory (ANL), Tech. Rep., 2009.

[3] L. A. Osadciw, Y. Yan, X. Ye, G. Benson, and E. White, "Wind turbine diagnostics based on power curve using particle swarm optimization," in *Wind Power Systems*. Springer, 2010, pp. 151–165.

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

[5] I. Sanchez, "Short-term prediction of wind energy production," *International Journal of Forecasting*, vol. 22, no. 1, pp. 43–56, 2006.

[6] C. C. Taylor, "Automatic bandwidth selection for circular density estimation," *Computational Statistics & Data Analysis*, vol. 52, no. 7, pp. 3493–3500, 2008.

[7] G. Lee, Y. Ding, M. G. Genton, and L. Xie, "Power curve estimation with multivariate environmental factors for inland and offshore wind farms," *Journal of the American Statistical Association*, vol. 110, no. 509, pp. 56–67, 2015.

[8] Y. Ding, J. Tang, and J. Z. Huang, "Data analytics methods for wind energy applications," in *ASME Turbo Expo 2015: Turbine Technical Conference and Exposition*. American Society of Mechanical Engineers, 2015, pp. V009T46A020–V009T46A020.

[9] A. Tindal, C. Johnson, M. LeBlanc, K. Harman, E. Rareshide, and A. Graves, "Site-specific adjustments to wind turbine power curves," in *AWEA Wind Power Conference, Houston*, 2008.

[10] H. Hwangbo, A. L. Johnson, and Y. Ding, "Spline model for wake effect analysis: Characteristics of a single wake and its impacts on wind turbine power generation," *IISE Transactions*, vol. 50, no. 2, pp. 112–125, 2018.

[11] B. Manobel, F. Sehnke, J. A. Lazzús, I. Salfate, M. Felder, and S. Montecinos, "Wind turbine power curve modeling based on gaussian processes and artificial neural networks," *Renewable Energy*, vol. 125, pp. 1015–1020, 2018.

[12] A. Marvuglia and A. Messineo, "Monitoring of wind farms power curves using machine learning techniques," *Applied Energy*, vol. 98, pp. 574–583, 2012.

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

[14] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[15] A. Goldstein, A. Kapelner, J. Bleich, and E. Pitkin, "Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation," *Journal of Computational and Graphical Statistics*, vol. 24, no. 1, pp. 44–65, 2015.