ICIC Express Letters
Part B: Applications
Volume **11**, Number **10**, October **2020**

ICIC International ©2020 ISSN 2185-2766
pp. 995–1000

A SHORT REVIEW ON PREDICTIONS FOR WIND POWER GENERATION – ITS LIMITATION AND FUTURE DIRECTIONS

Min Kyu Sim[1] and Jae-Yoon Jung[2,*]

[1]Department of Industrial and Systems Engineering
Seoul National University of Science and Technology
232, Gongneung-ro, Nowon-gu, Seoul 01811, Korea
mksim@seoultech.ac.kr

[2]Department of Industrial and Management Systems Engineering
Kyung Hee University
1732, Deogyeong-daero, Giheung-gu, Yongin-si, Gyeonggi-do 17104, Korea
*Corresponding author: jyjung@khu.ac.kr

Received February 2020; accepted May 2020

Abstract. *Controlling and predicting power generation are essential elements for efficient smart grid operations. In the case of wind energy, accurate prediction is even more important because humans have no control over its energy source. This study first reviews the line of prediction studies. Then, we generate point estimates using several machine learning methods and compare the performance and the pattern of the generated forecasts. We discuss the limitations of the point estimates and why probabilistic predictions are desirable. We suggest a few considerations for future studies.*
Keywords: Wind energy, Wind power generation, Point estimates, Machine learning, Numeric weather prediction

1. **Introduction.** Wind power generation accounts for 4.4% of global electricity generation in 2017 [1] and 14% of European Union's electricity generation in 2018 [2]. Since wind power generation does not emit pollutants, its importance and popularity are continuously growing. Control of power generation is a desirable element for efficient smart grid operations, but humans cannot control the energy source of wind power generation. The amount of wind power generation depends on various climate factors such as wind speed, temperature, atmosphere pressure, and the height of power plant. Thus, many studies have been conducted to predict the amount of wind power generation using these factors as predictor variables.

Studies on the prediction of wind energy are classified under the following several criteria. First, the prediction target variable can be either wind speed or wind power generation. Wind speed predictions for short prediction horizon may benefit power plant operations such as turbine configuration and maintenance. On the other hand, predictions on power generation take up the larger portion [3] and are more directly beneficial to electricity production. For this reason, this study focuses on the prediction problem for power generation.

For other criteria, prediction time horizon can be classified into very short-term (– 30 min), short-term (30 min – 6 h), medium-term (6 h – 24 h), long-term (24 h – 72 h), and very long-term (72 h –) [4]. In general, predictions of shorter time horizon can be used for immediate operations such as turbine configuration. Predictions of longer horizon can be used for plant installation planning and maintenance scheduling. Predictions of medium time horizon can be used for production planning and electricity trading.

DOI: 10.24507/icicelb.11.10.995

Studies can be also classified by the approach they adopt. Physical models aim to predict wind speed using the landscape of power plants and the movements of the surrounding clouds. When it comes to predicting climate variables using physical models, meteorologists in weather agencies routinely adopt physical models to generate numerical weather predictions (NWPs) that include the speed and direction of wind. Thus, the most popular approach for predicting wind power generation is the statistical models that apply statistical forecasting methods to NWPs. Since these models mostly utilize NWPs that are outputs from physical models, they are often called hybrid models. This study also belongs to these.

Statistical models can be further categorized into conventional models and modern learning-based models. Conventional models include time-series approaches such as auto-regressive (AR) [5], vector auto-regressive (VAR) [5], auto-regressive moving average (ARMA) [6], auto-regressive integrated moving average (ARIMA) [7], and error correction [8]. Learning-based models include support vector machine (SVM), boosting tree, random forest (RF), k-nearest neighbor (k-NN) [9]. A subset of learning-based approaches adopts neural network methods, which include artificial neural network (ANN), multi-layer perceptron network (MLP), and recurrent neural network (RNN) [10]. A survey paper [4] provides an extensive list for the main methods of relevant studies.

Most studies above are aimed to generate a point estimate, i.e., a single value that serves as a "best guess" or "best estimate" for the quantity of wind power generation. Point estimate is the most intuitive form of prediction, and various parametric approaches including the ones discussed above are applicable.

This study generates point estimates for wind power generation using several machine learning methods. This study compares the point estimates of machine learning methods in terms of their performance and the pattern of the forecasts. In our knowledge, no study has compared methods under the focus of their similar performance. Then, this study uniquely discusses the inherent limitation of point estimates and offers some guidelines for future studies.

This paper is organized as follows. In Section 2, problems are defined and relevant data sources are introduced. In Section 3, adopted machine learning methods are described and criteria for performance evaluation are specified. In Section 4, main results are presented. In Section 5, this paper concludes along with suggestions on future studies.

2. **Problem Statement and Preliminaries.** This study generates point estimates for wind power generation using several machine learning methods. After assessing the performances of prediction, this paper suggests the direction of future studies. For the experiment, this study uses historical power plant data and NWPs for the location. The Hankyung Wind Power Plant was built in 2004 in the Jeju Island, Korea. Each of its four units can produce 1500 kW of electricity per hour. The operating company, Korea Southern Power Co., Ltd., provided hourly power generation data from 2014 to 2017. A government-run weather agency (The Korea Meteorological Administration, KMA) publicly provides historic NWPs that include various climate variables. Data from nearby weather stations are collected for the same period (2014-2017). KMA announces weather forecasts in every three hours, and the look-ahead times for the forecasts are 4, 7, 10, 13, . . . , 67 hours. This study merges the two datasets and sets up a 4-hour ahead prediction problem. Under this problem setting, this study aims to identify the characteristics of point estimates. Among many climate variables in NWPs, we screen the variables that may be relevant to wind power generation.

It is known that an amount of wind power generation can be summarized in (1).

$$P = 0.5kC_p\rho AV^3, \tag{1}$$

where P is the amount of power output, k is a unit conversion constant, C_p is a dimensionless coefficient for the maximum power, ρ is air density, A is rotor swept area, and V is a wind speed. Among these factors, air density and wind speed are weather-related variables. Air density is determined by altitude, atmospheric pressure, temperature, and humidity. Thus, the following variables in the collected NWP data are used for this study: wind speed, temperature, and humidity.

3. **Methods.** This section briefly describes the machine learning methods applied in this study. These methods are popular supervised learning algorithms in the research line [11]. Linear regression (LR) is a classic modeling technique that seeks a linear relationship between predictors and a target variable. Support vector regression (SVR) is a variant of LR where small prediction errors are ignored in order to minimize the effect of outliers. Decision tree (DT) method splits the entire dataset into two subsets by searching for the best split condition that reduces the total error cost. This binary partitioning occurs recursively until a further partitioning is not beneficial. Two possible ensemble approaches can be added to the decision tree method. The bagging ensemble approach applied to the decision tree method leads to a method called random forest (RF). In RF, decision trees built from bootstrapped resampling construct a "forest", in which the prediction value is an average of values that each component tree produces. For a boosting ensemble approach, this study also utilizes an adaptive boosting (AdaBoost) method to the decision tree method. By focusing more on instances with larger error, the method boosts weak learners to produce powerful "committees". k-NN is a non-parametric method used for classification and regression. For each instance, the predicted value is based on the weighted average value of the k neighborhood instances where each of the weights is given as an inverse value of the distance between the target instance and the k nearest neighbors. ANN is a popular method in these days for classification and nonlinear regression problems. In ANN, multi-layered hidden nodes connect input and output nodes and the weights of nodes are optimized through the error backpropagation algorithm. This study applies these machine learning methods for predicting wind power generation.

Several criteria are available for evaluating point estimate. The review paper [4] reports that most studies use mean squared error (MSE), mean absolute error (MAE), and R-squared (R^2). The formulas of these error metrics are given in (2) to (4).

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y}_i)^2 = \mathrm{RMSE}^2, \tag{2}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \widehat{y}_i|, \tag{3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y}_i)^2}, \tag{4}$$

where y_i is the i-th actual value, $\widehat{y}_i$ is the predicted value for y_i, $\overline{y}_i$ is the mean of y_i, N is the number of predictions, and RMSE (root of MSE) implies the square root of MSE. Roughly speaking, MSE is mathematically convenient, while MAE is easier to interpret.

4. **Results.** This section generates point estimates for wind power generation using NWP. Several machine learning methods are utilized. The data for four years are split to a train set (3 years; years from 2014 to 2016) and a test set (1 year; year of 2017). Using the train set, five-fold cross-validation is performed with the random search algorithm. In optimizing hyper-parameters, the MSE is employed.

Table 1 presents the error metrics of the point estimates generated using several machine learning methods. ANN and RF exhibit the best performance in terms of RMSE. Overall

TABLE 1. Performance comparison

Method	R^2	RMSE	MAE	Maximum error
ANN	86.3%	200.6	133.2	1,194
RF	86.2%	200.6	131.3	1,197
Adaboost	86.0%	207.8	150.3	1,156
LR	85.7%	204.8	136.9	1,168
SVR	85.0%	211.6	133.0	1,273
DT	84.5%	212.2	139.9	1,263
k-NN	82.8%	225.2	151.6	1,347

R^2 of ANN is slightly higher. Though there are some differences, the performances are similar in the tested methods; the R^2 of the top four methods are ranged at the small interval of 85.7-86.3%.

Figures 1 and 2 present time-series plots for the forecast error. Figure 1 is for the last 14 days of the test set, and Figure 2 is for the entire test set. Both figures exhibit the fact that the generated forecasts are very similar in the several tested methods.

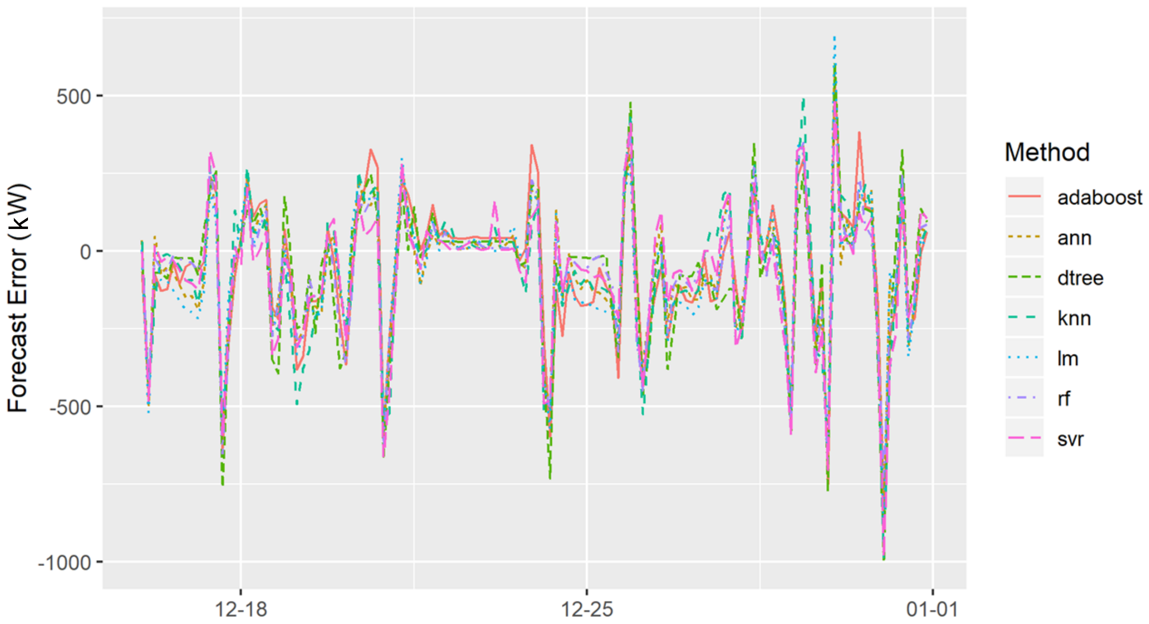

FIGURE 1. Forecast error for the last 14 days

Not only the performance (Figure 1) and the patterns (Figure 2 and Table 2) are similar, Table 2 confirms that the cross-correlation of the time-series is very high as well.

Given the same dataset and experiment design, the results of prediction are very similar even though the different algorithms are applied to building the models. This suggests the following points.

- Unless the scope of dataset is significantly widened, it is very difficult to enhance prediction accuracy for wind power generation.
- However, the dataset used in the above experiment contains most of the available data sources (NWPs and recent power generation) already examined by many previous studies.
- Thus, this striking similarity indicates the inherent limitation of the predictability on the amount of wind power generation.

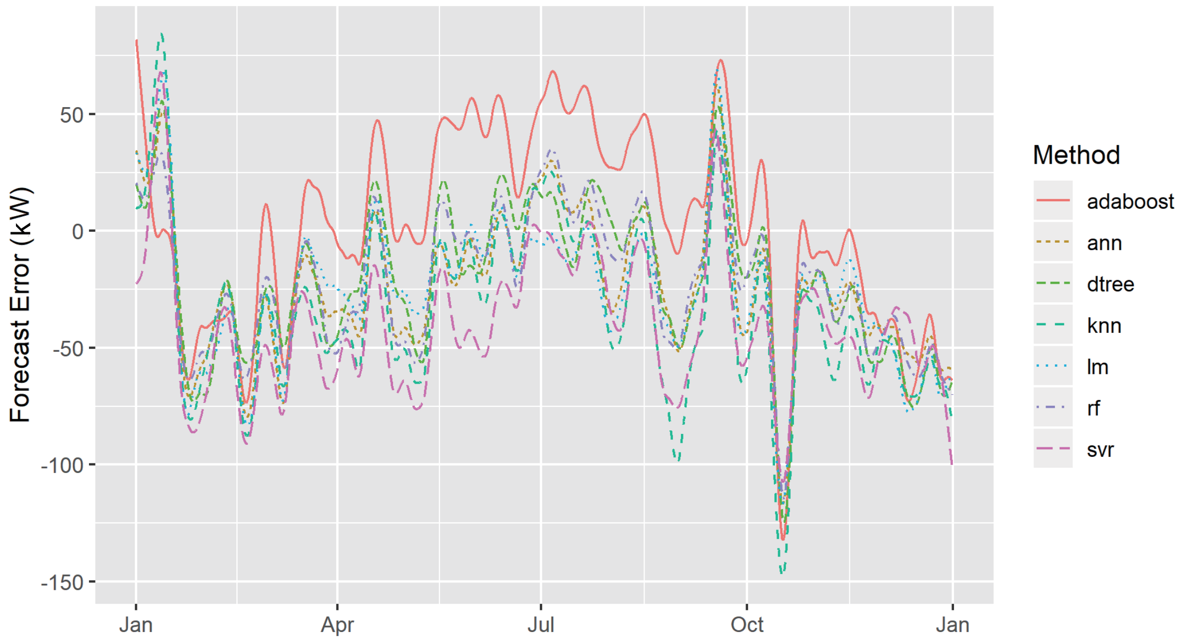

FIGURE 2. Forecast error for the entire test set (year of 2017) (Spline smoothing with degree of freedom equal to 50 is applied [12].)

TABLE 2. Cross-correlation of the forecast error (year of 2017)

	LR	RF	SVR	k-NN	DT	Adaboost	ANN
LR	1	0.95	0.94	0.90	0.91	0.93	0.99
RF		1	0.95	0.91	0.95	0.95	0.96
SVR			1	0.93	0.91	0.89	0.96
k-NN				1	0.88	0.89	0.92
DT					1	0.93	0.92
Adaboost						1	0.93
ANN							1

5. **Conclusion and Future Direction.** This study investigates and tests the predictability of wind power generation. Based on the results of literature survey, the predictors include NWPs and the recent power generation. The target variable is the amount of 4-hour ahead power generation. Our experiments employ several popular machine learning methods. As a result, it is found that the point estimates are very similar in terms of performance and patterns, regardless of the adopted machine learning algorithms. This leads us to conclude that point estimation for wind power generation has already reached its limit of improvement.

Emerged from the earlier studies that focused on point estimates, recent studies are evolved to 1) investigate sudden increases/decreases in the wind speed that degrade the accuracy of predictions [13] or 2) generate the interval estimate of power generation instead of point estimates [14,15]. The first group of the studies focuses on the phenomenon of sudden changes in wind speed, called as *wind ramp*. Wind ramp is generally defined in a binary form and efforts are made to identify climate conditions that cause ramps. Though the binary definition can be easily understood, mixing it up with parametric information can be quite tricky. The second group of the studies adopts statistical techniques, including bootstrapping methods in order to generate interval estimates. The remaining challenge is to include the findings in the studies of wind ramp when generating interval estimates. The interval output is potentially beneficial to decision makers a lot more than point estimates. We suggest that future studies would consider the following aspects.

- It is necessary to identify the conditions where the error of point estimate becomes large or small.
- NWPs and recent power generation must be utilized when generating interval estimates.
- Interval estimates should be generated according to different confidence levels.

Acknowledgment. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2019R1I1A1A0106148 2). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] *BP Statistical Review of World Energy 2018*, 67th Edition, BP p.l.c., London, UK, https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2018-full-report.pdf, 2018.

[2] C. Walsh and I. Pineda, *Wind Energy in Europe in 2018 – Trends and Statistics*, Wind Europe, Brussels, Belgium, https://windeurope.org/wp-content/uploads/files/about-wind/statistics/WindEurope-Annual-Statistics-2018.pdf, 2019.

[3] I. Colak, S. Sagiroglu and M. Yesilbudak, Data mining and wind power prediction: A literature review, *Renewable Energy*, vol.42, pp.241-247, DOI: 10.1016/j.renene.2012.02.015, 2012.

[4] I. Okumus and A. Dinler, Current status of wind energy forecasting and a hybrid method for hourly predictions, *Energy Conversion and Management*, vol.123, pp.362-371, DOI: 10.1016/j.enconman.2016.06.053, 2016.

[5] C. Croonenbroeck and D. Ambach, A selection of time series models for short- to medium-term wind power forecasting, *Journal of Wind Engineering and Industrial Aerodynamics*, vol.136, pp.201-210, DOI: 10.1016/j.jweia.2014.11.014, 2015.

[6] T. Senjyu, A. Yona, N. Urasaki and T. Funabashi, Application of recurrent neural network to long-term-ahead generating power forecasting for wind power generator, *Proc. of IEEE PES Power Systems Conference and Exposition*, Atlanta, GA, USA, pp.1260-1265, DOI: 10.1109/PSCE.2006.296487, 2006.

[7] J. Wang and J. Hu, A robust combination approach for short-term wind speed forecasting and analysis – Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts using a GPR (Gaussian Process Regression) model, *Energy*, vol.93, pp.41-56, DOI: 10.1016/j.energy.2015.08.045, 2015.

[8] L. Wang, X. Li and Y. Bai, Short-term wind speed prediction using an extreme learning machine model with error correction, *Energy Conversion and Management*, vol.162, pp.239-250, DOI: 10.1016/j.enconman.2018.02.015, 2018.

[9] A. Kusiak and Z. Zhang, Short-horizon prediction of wind power: A data-driven approach, *IEEE Trans. Energy Conversion*, vol.25, no.4, pp.1112-1122, DOI: 10.1109/TEC.2010.2043436, 2010.

[10] J. Chen, G. Q. Zeng, W. Zhou, W. Du and K. D. Lu, Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization, *Energy Conversion and Management*, vol.165, pp.681-695, DOI: 10.1016/j.enconman.2018.03.098, 2018.

[11] A. M. Foley, P. G. Leahy, A. Marvuglia and E. J. McKeogh, Current methods and advances in forecasting of wind power generation, *Renewable Energy*, vol.37, no.1, pp.1-8, DOI: https://doi.org/10.1016/j.renene.2011.05.033, 2012.

[12] T. J. Hastie, Generalized additive models, in *Statistical Models in S*, Routledge, 2017.

[13] C. Gallego-Castillo, A. Cuerva-Tejero and O. Lopez-Garcia, A review on the recent history of wind power ramp forecasting, *Renewable and Sustainable Energy Reviews*, vol.52, pp.1148-1157, DOI: 10.1016/j.rser.2015.07.154, 2015.

[14] A. Khosravi, S. Nahavandi and D. Creighton, Prediction intervals for short-term wind farm power generation forecasts, *IEEE Trans. Sustainable Energy*, vol.4, no.3, pp.602-610, DOI: 10.1109/TSTE.2012.2232944, 2013.

[15] R. Li and Y. Jin, A wind speed interval prediction system based on multi-objective optimization for machine learning method, *Applied Energy*, vol.228, pp.2207-2220, DOI: 10.1016/j.apenergy.2018.07.032, 2018.