J Electr Eng Technol.2017; 12(6): 2177-2186
http://doi.org/10.5370/JEET.2017.12.6.2177

ISSN(Print) 1975-0102
ISSN(Online) 2093-7423

Very Short-Term Wind Power Ensemble Forecasting without Numerical Weather Prediction through the Predictor Design

Duehee Lee*, Yong-Gi Park, Jong-Bae Park* and Jae Hyung Roh†**

Abstract – The goal of this paper is to provide the specific forecasting steps and to explain how to design the forecasting architecture and training data sets to forecast very short-term wind power when the numerical weather prediction (NWP) is unavailable, and when the sampling periods of the wind power and training data are different. We forecast the very short-term wind power every 15 minutes starting two hours after receiving the most recent measurements up to 40 hours for a total of 38 hours, without using the NWP data but using the historical weather data. Generally, the NWP works as a predictor and can be converted to wind power forecasts through machine learning-based forecasting algorithms. Without the NWP, we can still build the predictor by shifting the historical weather data and apply the machine learning-based algorithms to the shifted weather data. In this process, the sampling intervals of the weather and wind power data are unified. To verify our approaches, we participated in the 2017 wind power forecasting competition held by the European Energy Market conference and ranked sixth. We have shown that the wind power can be accurately forecasted through the data shifting although the NWP is unavailable.

Keywords: Wind power forecasting, Ensemble forecasting, Gradient boosting machine

1. Introduction

Forecasting begins from estimating the pattern between historical predictors and targets and ends by applying the pattern to new predictors. The predictor can be historical measurements or forecasted predictors. In wind power forecasting, an example of the former case is to forecast an hour-ahead wind power based on the historical wind power measurement, and an example of the latter case is to forecast an hour-ahead wind power based on the wind speed that is predicted to blow after an hour. Since there is no forecasted predictor in the former case, the latter case has generally shown higher performance [1].

The goal of the latter case is to convert predicted meteorological weather conditions to wind power outputs. The predicted weather data is called the numerical weather prediction (NWP), which includes predicted temperature, radiation, precipitation, wind speed, and so on. The representative NWP generation models are the weather research and forecasting (WRF) model, mesoscale model 5 (mm5), and European Centre for Medium-Range Weather Forecasts (ECMWF) [2]. The NWP can be estimated on super computers by solving the partial differential equations (PDE), in which current weather conditions are used as initial states. The PDEs describe the movement of flows, such as air or temperature, from now to the target future, so if we know the initial state of weather conditions, we can make a conjecture about the future state of weather conditions [3].

However, practically, it is difficult to collect the historical NWP for the old target data and new NWP for the new target data since many computational resources are required to estimate the many weather variables in the NWP. Besides, the NWP is not sampled at the same sampling interval with wind power.

When the NWP has not been available, the time series models have been used [4]. In this case, since we do not have any evidence about the future, the only available data is the measured weather conditions and wind power. Therefore, the wind power is recursively forecasted in the time series models based on the historical wind power or the forecasted wind power in previous steps. The weakness of the time series models is that they have limited computational power to analyze the nonlinear characteristics between wind speed and wind power since the time series models are based on linear algorithms.

Recently, machine learning-based advanced forecasting algorithms have been developed, and they perform better than time series models [5]. The generalized linear model, gradient boosting machine (GBM), and the random forest (RF) are representative machine learning-based forecasting models. These models are specialized to extract the nonlinear relationship between the forecasted predictors and targets. Although the NWP is not provided as the forecasted predictor in our study, we can utilize these advanced forecasting algorithms by shifting the historical

† Corresponding Author: Dept. of Electrical Engineering, Konkuk University, Korea. (jhroh@konkuk.ac.kr)
* Dept. of Electrical Engineering, Konkuk University, Korea. ({hello.Keanu, jbaepark}@konkuk.ac.kr)
** Dept. of Electrical and Electronic Engineering, Youngsan University, Korea. (draco.park@ysu.ac.kr)
Received: May 15, 2017; Accepted: August 5, 2017

weather and wind power data.

In this study, we propose a way to shift the weather conditions and wind power and generate predictors to forecast the very short-term wind power through the machine learning-based forecasting algorithms without using the NWP. In shifting the training data, which consists of the historical weather conditions and wind power outputs, we shift the training data to the future and match the most recently sampled training data to the target prediction horizon. Then, we analyze the relationship between the target wind power and previously measured weather conditions and wind power outputs, which were measured before the target wind power is measured. Therefore, we forecast the future wind power based on the previously measured training data, not on the concurrently measured training data. Furthermore, at every target prediction horizon, we build a different forecasting machine to reflect the forecasting performance deterioration as the target time increases. In addition, we change the sampling interval of wind power and weather data to find the optimal sampling interval.

In order to verify our forecasting algorithms, we participated in the 2017 forecasting competition operated by the European Energy Market Conference, and we ranked 6th among 42 participants [6]. In this study, the performance of our advanced training data restructuring is compared to the performance of classical time series models to justify the data shifting.

The ultimate goal of this paper is to provide the specific forecasting steps and explain how to design the forecasting architecture and training data sets to forecast the very short-term wind power when the NWP is unavailable, and when the sampling periods of the wind power and NWP are different. The key contributions of this research are as follows: i) We suggest the novel training data structure to implement the machine learning techniques without the NWP. In this frame, users can implement whichever tools and functions they need according to the situations ii) We implement the temporal hierarchical forecasting scheme to reflect the deteriorating forecasting performance as the time goes so that users can forecast wind power accurately without the NWP. iii) We suggest the optimal sampling interval when the wind power and weather information have different sampling intervals. iv) Finally, we suggest the ensemble forecasting algorithm of various machine learning algorithms in order to compensate the low performance of classical time series models when the NWD is unavailable.

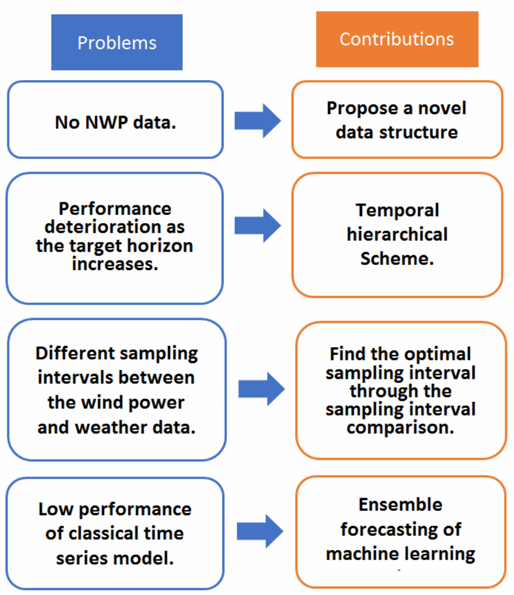

Fig. 1. Problems and our solutions that we contribute

Problems and our solutions that we contribute to solve the problems are enumerated in Fig. 1.

The contents of this paper are as follows. In 2. Section, the forecasting data is introduced. Then, the basic forecasting structure and algorithms are described. 3. Section describes how to design the training data, which will be used in 2. Section, to maximize the forecasting performance. In 4. Section, the forecasting performance and results from benchmarks are introduced. Furthermore, the forecasting result is analyzed. Finally, 5. Section summarizes the results.

2. Forecasting Data & Techniques

In this section, the forecasting data and the forecasting architecture are described. Furthermore, machine learning-based forecasting algorithms and time series models are introduced.

2.1 Forecasting data

The forecasting data from 2016/01/01 to 2017/04/16 is provided. For the given forecasting data, we forecasted 152 wind power outputs every 15 minutes for 38 hours from 10 a.m. to the next day 0 a.m. To mirror real life conditions, the new measured wind power and weather data during the previous 24 hours was provided every day at 8:00.

In other words, at 8 a.m. at every day, the historical wind power data and weather data between 8 a.m. on the day before and 8 a.m. on that day was provided. Therefore, we had to forecast the wind power for the next 38 hours within two hours by 10 a.m. if we wanted to utilize the most recently measured wind power and weather data.

The data used in this study consisted of different types and was sampled on various intervals. The wind power was sampled every 15 minutes, but the measured weather conditions was sampled every three hours. Since the sampling interval of wind power and weather information was different, the proper sampling interval had to be determined. The wind power and weather data were measured from 2016/01/01 to 2017/04/03, and from 2017/04/04, the daily new wind power and weather data were provided every day at 08:00. The weather data consisted of u-axis wind speed, v-axis wind speed, temperature, and global radiation. These variables were measured from 10 different weather stations. The training data was built by collecting weather data from all 10 weather stations.

The forecasting error is measured by the mean absolute error (MAE), which is defined as

$$\text{MAE} = \frac{\sum_{i=1}^{N} |y_i - \hat{y}_i|}{N} \tag{1}$$

where the $\hat{y}_i$ is predicted value, y_i is the actual value, and N is the number of predictors. In this study, the forecasting error is also measured by the mean absolute percentage error (MAPE), which is defined as

$$\text{MAPE} = \frac{100}{N} \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right|. \tag{2}$$

2.2 Training data transformation

The first key point of the training data shift is to generate the predictor for the forecast at the target prediction horizon. We shift the weather data a few hours ahead so that the most recent weather data corresponds to the target prediction horizon. This shift assumes that the past weather information will affect the future wind power. At the final target time, we want to match the final weather information to the final target prediction horizon, which is the next day at 23:00. Table 1 shows the structure of the forecasting training data when the target prediction horizon is 2017/04/05 20:00. Six a.m. is the last hour when the wind power and weather data are measured simultaneously. The wind power is also measured by 07:00, so there is no weather data after 07:00. Although we forecast wind power from 08:00 to maintain the continuity of the forecasts from hour to hour, only the wind power forecasts after 10:00 are submitted for the daily ranking.

Table 1. The training data structure

	Wind power	Weather data	1st past wind power	…	7th past wind power
Training data	2016/1/1 04:00	2016/1/1 00:00 (First observed weather)	2016/1/1 13:00	…	2016/1/1 06:00
	…	…	…	…	…
	2017/4/5 06:00	2017/4/4 16:00	2017/4/5 05:00	…	2017/4/4 22:00
	2017/4/5 07:00 (Last observed wind power)	2017/4/4 17:00	2017/4/5 06:00	…	2017/4/4 23:00
Testing data	2017/4/5 08:00	2017/4/4 18:00	2017/4/5 07:00	…	2017/4/5 00:00
	…	…	…	…	…
	2017/4/5 20:00	2017/4/5 06:00 (Last observed weather)	2017/4/5 19:00	…	2017/4/5 12:00

For example, in order to forecast the wind power at 20:00 on 2017/4/5, we use the weather data that was measured at 06:00 on 2017/4/5. Moreover, the forecasted wind power outputs between 12:00 on 2017/4/5 and 19:00 on 2017/4/5 are also used to forecast wind power at 20:00 on 2017/4/5. Therefore, in order to build the training data set, we shift the weather data 14 hours ahead.

The second key point of the training data shift is to have individual forecasting models for each target prediction horizon. Since the forecasting performance deteriorates as the target prediction horizon increases, in order to avoid the spread of performance deterioration, we build individual forecasting machines.

For example, when the sampling interval is 15 minutes, we build 152 machines, and when the sampling interval is an hour, we build 38 machines. Each forecasting machine has its own training data. We can connect these forecasting machines by delivering the forecast of the previous target-time model to the next target-time model. The advantage of this method is that the forecasting machines corresponding to the early target times can forecast wind power outputs with very high accuracy. The disadvantage is that the forecasting error can be propagated since forecasting machines corresponding to later target hours should forecast the wind power based on the previous forecasts. However, the effect of the error propagation might not worsen the forecasting performance seriously, since there are many other variables.

2.3 Training data separation

In order to represent the continuous nature of wind flow in the training data, we do not separate the forecasting model for each target prediction horizon into 24 different forecasting models for every hour. Although the wind power has a strong daily pattern because of the daily rotation of the Earth [7], if we use the advanced forecasting models based on the machine learning algorithms, we can forecast the wind power by automatically classifying the training data for each hour. Besides, since training data in this study, spans only a year and four months or four months, if we split the forecasting data into 24 sets, each model will not have enough training data. Therefore, we use all the training data and do not separate the model and data into 24 models.

2.4 Sampling interval selection

We find the optimal sampling interval and sample the wind power and weather data with the optimal sampling interval. The wind power was originally sampled every 15 minutes, and the weather data was originally sampled at every three hours. We select the best sampling interval by

Table 2. Performance comparison of sampling intervals

Sampling interval	15 min	1 hour	3 hour
MAE	22.2345	23.5494	31.4159
MAPE	203.2352	214.315	340.6912
Time (seconds)	3842.2412	312.4902	242.3415

comparing the forecasting performances of three sampling intervals: 15 minutes, one hour, and three hours. Although in this study one hour is only tested as a new sampling interval, other sampling intervals that are multiple of the smallest sampling intervals can be tested. Since the data sampled at the smallest sampling interval has the highest data resolution, the smallest sampling interval should be the minimum one of candidates.

First, we interpolate the weather data every 15 minutes. Second, we down sample the wind power every three hours. Third, we interpolate the weather data at every hour, and we down sample the wind power at every hour. The first approach provides the most information, but it takes the longest time to compute, and would be difficult to complete within the allotted two hours. Besides, the interpolation of the weather data can contaminate the data. For the third approach, if we down sample the wind power every three hours, we will lose a lot of information about wind power. Therefore, the second approach is the best option, but we choose the optimal sampling interval by comparing their forecasting performances. The performance between sampling intervals are enumerated in Table 2. The final performance of forecasting algorithm is calculated by averaging the MAEs or MAPEs of wind power forecasts for two weeks.

Although the performance of the three-hour sampling period has the best performance, because of the computational time, we select the one hour sampling period. Therefore, when there are two different sampling intervals, we recommend testing multiple values of the smaller sampling interval and selecting the best performing sampling interval. However, further research should be performed to develop a systematic approach to find an optimal interval. Finally, we down sample the wind power at every hour, and we interpolate the historical weather data at every hour. Furthermore, the hourly forecasted wind power should be interpolated at every 15 minutes to submit the 15-min sampled forecasts.

2.5 Auxiliary forecasting techniques

In addition to training data shift, selecting the optimal sampling interval, and utilizing the data interpolation, we also considered two forecasting techniques to increase the forecasting performance: the seasonal pattern management and outlier detection. The seasonal pattern management was used to extract the daily pattern in the time series models under the assumption that there is only daily period in the seasonal patterns. We use the pattern management

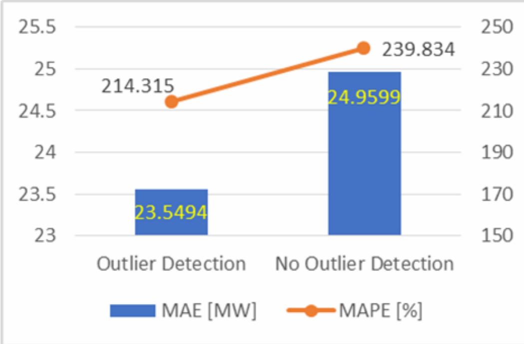

Fig. 2. The performance comparison between with or without outliers

technique differently according to forecasting algorithms. Since the time series models use only the wind power recursively, we can take out the daily pattern from wind power. On the contrary, in the machine learning-based algorithms, the seasonal pattern management was not necessary. If we extract the daily pattern from wind power, we should also extract the daily pattern from the training data. Otherwise, the deterministic daily pattern in the wind power might be confused with the fluctuating daily pattern in the training data.

Regarding the second forecasting technique, we extract outliers from training cases. The outliers are defined as training cases that have higher errors than other training cases. The outliers are detected through the Ridge regression, since the Ridge regression can estimate the forecasts very quickly. The Ridge regression is explained in [8]. It should be noted that we detect outliers based only on the forecasting performance of the training data, not on the forecasting performance of the testing data. The forecasting performance with and without the outliers detection is shown in Fig. 2. We can observe that removing outliers can slightly increase the forecasting performance.

3. Ensemble Forecasting of Forecasting Machines

In this section, the machine learning-based ensemble forecasting algorithms and time series models are described.

3.1 Data manipulation

Before explaining the forecasting machines, we introduce how the training and target data are processed and applied. We shift the four historical weather data, which consists of u-axis wind speed, v-axis wind speed, temperature, and global radiation. Since these variables were measured from 10 different weather stations, so there are 40 variables. Furthermore, we add time information: month, day of year, and hour. We also add the past wind power for 32 hours. Therefore, we use 75 variables as the training data. At the training time t, the historical training data is denoted by $\boldsymbol{X}_t$ and the target wind power is denoted by y_t. Furthermore, at the target time, the future training data is denoted by $\hat{\boldsymbol{X}}_t$, and the forecasted wind power is denoted by $\hat{y}_t$. The training data are normalized by subtracting the mean values and dividing by the standard deviations.

3.2 Ensemble forecasting algorithms

In the ensemble forecasting, the final prediction is calculated by melding the results from many individual forecasting machines into the final ensemble predictor. There are four different ensemble forecasting algorithms. The first and second algorithms can be used when different types of individual forecasting machines are used, and the third and fourth algorithms can be used when only a decision tree is used as an individual forecasting machine. First, when different types of forecasting machines are used in the ensemble forecasting, we can simply average the results of those machines. Second, with different types of forecasting machines, in this case, we can simply meld the results by weighted averaging, where all forecasts are averaged after multiplying weight factors. The forecasting machines with better performance are multiplied by higher weight factors so that the contribution of high performing forecasting machines increases linearly with weight factors.

The first algorithm is used when the performance of each forecasting machine cannot be measured through the cross validation, and the second algorithm is used when the performance of each forecasting machine can be measured through the cross validation. The performance of the second algorithm is higher than that of the first algorithm, but the second algorithm takes more time to run the cross validation, so the second algorithm is recommended when there is enough time to run the cross validation.

It should be noted that since there is not enough time to measure the performance through the cross validation, the forecasts from various forecasting machines are uniformly averaged with an equal weight by following the first algorithm.

The third algorithm is the RF, and the fourth algorithm is the GBM. Both third and fourth algorithms use a decision tree as an individual forecasting machine, so they are good at detecting non-linear relationships. The performances of both algorithms differ by applications, so users should choose an algorithm after measuring their performances. The RF and GBM are explained in detail in following subsections. In this study, various forecasting machines including the RF and GBM are used to increase the forecasting performance.

3.3 Random forest

In the RF, the results of many decision trees are simply averaged. However, the RF is different from the first ensemble algorithm, which simply averages forecasts of multiple individual forecasting machines. Each forecasting machine, which is a decision tree in the RF, has its own training data sets by randomly selecting training variables and cases. The performance of the RF can be different from the number of selected predictors, the number of cases, and the depth of the branches in the trees.

The process of building a decision tree is described. Building a decision tree means to divide the training data into J distinct and non-overlapping boxes, $R_1, \cdots, R_J$. The boxes should minimize the loss as

$$L = \sum_{i=1}^{J} \sum_{i \in R_j} \left(y_i - \hat{y}_{R_j} \right)^2 \tag{3}$$

where $\hat{y}_{R_j}$ is the mean response for training cases in the box. Considering all possible sets of boxes to find the optimal set of boxes is computationally infeasible. Therefore, we can use the recursive binary splitting method [9], which is also known as a greedy and top-down method. In the binary splitting method, each split is represented by two new branches further down. The prepositive split is determined not considering future coming splits, so it is called a greedy method. Furthermore, the split moves from the root to the leaves, so it is called a top-down method.

Then, we select the predictor X_j to split and the splitting point s in X_j so that the resulting tree will have the minimum loss. In this process, we should consider all predictors $X_1, \cdots, X_p$, where p is the number of predictors. In detail, for any j and s, we define the pair of half planes as

$$R_1(j,s) = \left\{ X \mid X_j < s \right\} \ \ and \ \ R_2(j,s) = \left\{ X \mid X_j \geq s \right\} \tag{4}$$

For the given two planes, we should find the j and s that minimize the sub-loss l as

$$l = \sum_{i:x_i \in R_1(j,s)} (y_i - \hat{y}_{R_1})^2 + \sum_{i:x_i \in R_2(j,s)} (y_i - \hat{y}_{R_2})^2 \tag{5}$$

where $\hat{y}_{R_1}$ is the mean response for the training cases in $R_1(j,s)$, and where $\hat{y}_{R_2}$ is the mean response for the training cases in $R_2(j,s)$. The j can be found by minimizing loss l in (5). For the fixed j, we can find the s by differentiating l. Next, we could perform this process repetitively to find the next splitting predictor and splitting point. However, instead of splitting the entire predictor, we should split the half plane of the selected predictor until each box, which is called a leaf, has a minimum number of observations. In this simulation, it is set at two. The result is calculated by averaging the response in the corresponding box. After averaging all weak trees with different training data, we could get the final response of the RF.

In the forecasting process, we can use the process mentioned below:

$$RF = RandomForest\left(\boldsymbol{X}, y\right).$$
$$\hat{y} = RF\left(\hat{\boldsymbol{X}}\right)$$

In this process, *RandomForest* is the software function, and the *RF* is the class about the random forest. After training the class *RF* with training weather $\boldsymbol{X}$ and wind power y, the class *RF* will have the relationship between $\boldsymbol{X}$ and y. Then, if we apply the class *RF* to the shifted training data $\hat{\boldsymbol{X}}$, we can get the forecasted wind power $\hat{y}$. It should be noted that the subscript about the time is omitted for simplicity.

3.4 Gradient boosting machine

On the contrary to the RF, in the GBM, the individual forecasting machine is updated at every iteration [10]. The goal of the GBM is to find the final function $F(\mathbf{x})$ that minimizes the expected value of loss function $L(y, \mathbf{x})$ for all training data $\boldsymbol{X}$ and target data y. The $F(\mathbf{x})$ that satisfy the goal can be found through the minimization process:

$$\begin{aligned} F(\mathbf{x}) &= \arg\min_{f(x)} \Phi(y, f(\mathbf{x})) \\ &= \arg\min_{f(\mathbf{x})} E_{\mathbf{x}}[E_y[L(y, f(\mathbf{x}))] \mid \mathbf{x}]. \end{aligned} \tag{6}$$

The first key idea in the GBM is to represent the $F(\mathbf{x})$ as the sum of all individual functions, $f(\mathbf{x})$, which is called a weak function. The $F(\mathbf{x})$ can be represented by $f(\mathbf{x})$ as

$$F(\mathbf{x}) = \sum_{m=1}^{M} f_m(\mathbf{x}) \tag{7}$$

The second key idea is to continuously upgrade the $F(\mathbf{x})$ by adding a new weak function $f(\mathbf{x})$ to the previous $F(\mathbf{x})$. The $f(\mathbf{x})$ can be represented as the gradient scale ρ_m and a gradient following tree function $h(\mathbf{x}, \boldsymbol{\theta})$:

$$f_m(\mathbf{x}) = -\rho_m h(\mathbf{x}, \boldsymbol{\theta}_m) \tag{8}$$

The parameter $\boldsymbol{\theta}_m$ in the $h(\mathbf{x}, \boldsymbol{\theta}_m)$ is searched so that the tree $h(\mathbf{x}, \boldsymbol{\theta}_m)$ is fitted to the pointwise gradient of the expected loss function $g_m(\mathbf{x}_i)$ of the most recent final function $F_{m-1}(\mathbf{x})$ at every $\mathbf{x}_i$, where $i \in \{1, \ldots, N\}$, and N is the number of all training data. The $\boldsymbol{\theta}_m$ can be estimated as

$$\boldsymbol{\theta}_m = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} (g_m(\mathbf{x}_i) - \beta h(\mathbf{x}_i, \boldsymbol{\theta}_m))^2 \tag{9}$$

By minimizing the sum of all errors between the gradient and a tree function, we can aggressively train the training data $\mathbf{x}_i$ of the worst performance. The $g_m(\mathbf{x}_i)$ is defined as

$$g_m(\mathbf{x}_i) = \left\{ \left[\frac{\partial \Phi(f(\mathbf{x}_i))}{\partial f(\mathbf{x}_i)} \right]_{f = f_{m-1}} \right\} \tag{10}$$

Furthermore, the ρ_m is defined when it minimizes the sum of all loses as

$$\rho_m = \arg\min_{\rho} \sum_{i=1}^{N} L\left[y_i, f_{m-1}(x_i) - \rho h(x_i, \theta_m)\right] \tag{11}$$

In summary, the GBM finds the $F(\mathbf{x})$ by fitting the tree function to the gradient of expected loss function and optimizing the scale of the tree function. It should be noted that when applying the GBM to the X and y, we can use the same process in the random forest with different function.

3.5 Time series models

In this subsection, we introduce the time series models, which are also well explained in [11]. The time series models have been widely used to forecast the wind power when the NWP is not available since they can recursively forecast future wind power outputs solely based on the historical wind power outputs. When we compare the time series models to the machine learning-based forecasting algorithms, the time series models also analyze the relationship between the past and future wind power outputs. However, the time series models recursively forecast the future wind power right at one prediction horizon ahead based on the past wind power or the forecasted wind power when the past wind power is not available as the prediction horizon proceeds. The basic assumption on this process is that the target variable is sampled sequentially. The time series models can also use the additional training variables, but the training data must exist at future prediction horizon. Therefore, since the NWP is not available in this study, in order to use the additional training variables, we should use the data shifting idea as machine learning-based algorithm do.

Two different time series models can be used in the wind power forecasting: the univariate time series model and univariate time series model with exogenous terms. There is also the multivariate time series model, but it is out of scope of this paper. First, the univariate time series model used in this study is the autoregressive (AR) model. It is defined as

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t \tag{12}$$

where the y_t is the wind power output, the ϕ_i is the i_{th}

parameter of y_{t-i} term, the p is the system order, and the ε_t is the white noise. Theoretically, the order p can be found through the AIC or BIC [12], and the parameters ϕ_i can be estimated through the Yule-Walker equations [13]. However, in this study, the p and ϕ_i are automatically estimated through the MATLAB System Identification Toolbox [14].

Second, the univariate time series model with exogenous terms used in this study is the AR with Exogenous (ARX) model. It is defined as

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{k=1}^{K} \sum_{i=1}^{q(k)} \pi_{t-1}^{k} U_{t-1}^{k} + \varepsilon_t \quad (13)$$

where the U_t is the exogenous predictor, and the π_t is the parameter of the U_{t-i} term. Furthermore, the k is the number of predictors, and the $q(k)$ is the order of k_{th} exogenous term. Although the NWP is not available, if we transform the training data, we can have exogenous predictor for all forecasting duration. Then, we can use the univariate time series with exogenous term. Parameters can also be estimated by the Yule-Walker equations, but in this study, they are estimated by the MATLAB System Identification Toolbox.

In summary, when the NWP is not available, we can use the univariate time series models without transforming the training data. With transforming the training data, we can use the univariate time series model with exogenous terms. In this study, we use the AR to measure the performance of the univariate time series model and the ARX to measure the performance of the univariate time series model with exogenous terms.

4. Simulation

In this section, we introduce our forecasting result and their benchmarks. Then, we present the performance of ensemble forecasting algorithm. Finally, we show that the forecasting performance deteriorates as the prediction horizon increases. This phenomenon supports that we use independent forecasting machines at every prediction horizon in order to stop spreading forecasting errors across prediction horizons.

Table 3. Comparison of performance benchmark

Ranking	User Name	MAE
1	P9	19.591
2	4C	20.507
3	P5	20.559
4	Zephyr	21.604
5	Dmlab	22.119
6	Our Team	22.521
7	Return42	22.633
8	P25	22.807
9	DSAP_XXQ	23.347
10	DSAP_group1	23.481

4.1 Forecasting benchmark

The average MAEs between 2017/04/03 and 2017/04/16 of 10 teams are enumerated in Table 3. We ranked 6th among 42 participants [16].

4.2 The forecasting result

We compare the performance of machine learning-based algorithms with the data shift to the performance of time series models without the data shift in order to justify the data shift. In Fig. 4. the performance of five different forecasting machines and the performance of their average are plotted. The performance is the averaged one for two weeks. We can see that the machine learning-based forecasting algorithms, such as RF and GBM, have a higher performance than the time series models, such as AR and ARX. Therefore, the shift of the weather data is necessary to increase the forecasting performance. The machine learning-based forecasting algorithms analyze and use the pattern between the wind power and the weather data, which was measured before the wind power was measured.

Furthermore, the performance of the ensemble forecasting,

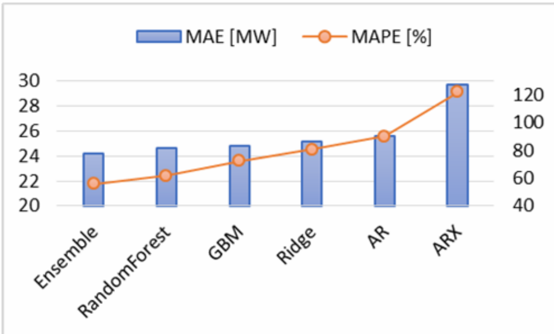

Fig. 3. The performance of ensemble forecasting

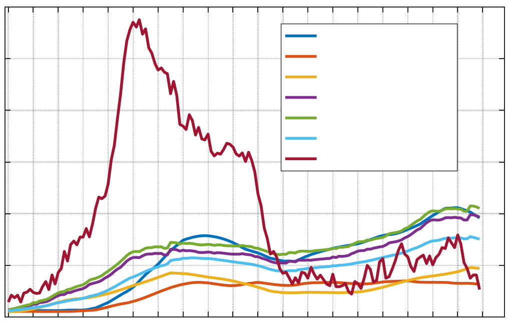

Fig. 4. The forecasts of five forecasting machines and their ensemble forecasting machine

where the forecasts of all forecasting machines are averaged without weight factors, is analyzed. We can observe that the ensemble of results of all forecasting machines has the lowest forecasting error. Moreover, the performance of the ARX is worst, since if the number of predictors is too large, the ARX is easily affected by noises in the predictors, and the parameter estimation algorithm in the ARX does not perform well if the number of parameters increases as the number of predictors increases.

Moreover, the forecasts of five forecasting machines and ensemble forecasting are plotted in Fig. 4. We forecast the wind power at every hour between 10:00 on 2017/04/12 and 23:45 on 2017/04/13 and then interpolate forecasts at every 15 minutes.

In Fig. 4., we observe that the machine learning-based forecasting can respond to the change of historical weather data actively although the forecasts of time series models follow the actual wind power for early prediction horizons.

The red line represents the actual wind power, and the cyan line represents the wind power that is forecasted by the ensemble forecasting algorithm, where the Ridge regression, RF, ARX(8), and GBM algorithms are used. Furthermore, the blue line represents the result of the AR(8) model. We can observe that the AR model follows the actual wind power at the early target times, but it fails to follow the wind power after 22:00, since it does not use the trends from the historical weather data. In contrast, the forecasts of the ensemble forecasting algorithm follow the actual wind power well up to 20:00, but after 00:00, it goes back to the average of the historical wind power. It shows that the best policy is to forecast the mean value when the relationship between predictors and target values is not strong.

4.3 Hourly forecasting performance

In this section, we measure the hourly forecasting performance as the target prediction horizon increases. The forecasting performances deteriorates as the prediction horizon is far from the time when the last wind power is observed since the dependence of forecasted value on measured weather is getter worse, so it becomes harder to forecast the wind power accurately.

In Fig. 5, the average forecasting errors between 2017/4/3 and 2017/4/16 are plotted with respect to the target prediction horizon. The forecasting errors increase overall as the target prediction horizon increases. It supports the reason why we build individual forecasting machines for each prediction horizon. However, the forecasting errors between 04:00 and 06:00 do not follow the overall non-decreasing trend. We set a hypothesis that the increased forecasting performance as a daily pattern decreases around that time.

The hourly daily pattern of wind power is plotted in Fig. 6.

In Fig. 6, we can observe that the wind power generally decreases between 8 AM and 14 AM. However, there is a slight time shift between the decreasing time of forecasting errors and the decreasing time of daily pattern. In order to explain this discrepancy, in Fig. 7., we plot the hourly averaged wind power of actual wind power between 2017/4/3 and 2017/4/16, which corresponds to the forecasting duration. We can also see that the solution wind power also decreases between 4 hour and 8 hour. That is why the average forecasting error does not follow the non-decreasing function. Furthermore, since the prediction of

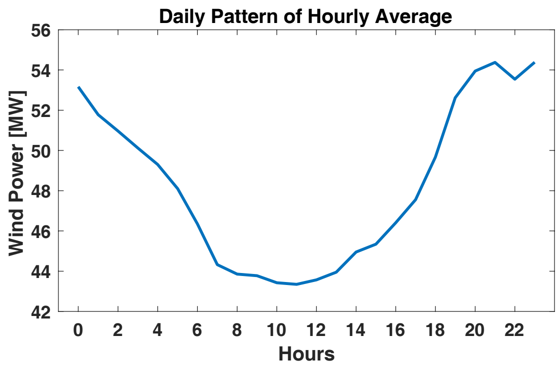

Fig. 6. The daily pattern of wind power is sampled at every hour

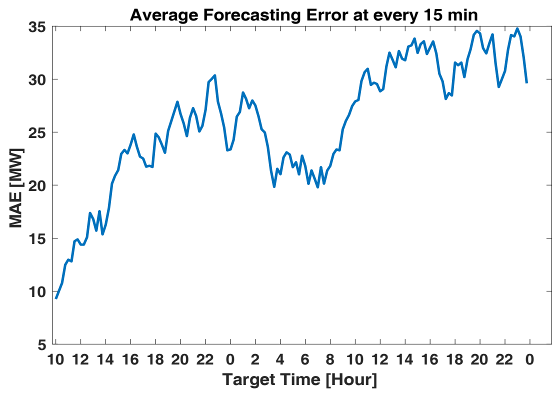

Fig. 5. Average forecasting errors with respect to the target forecasting time

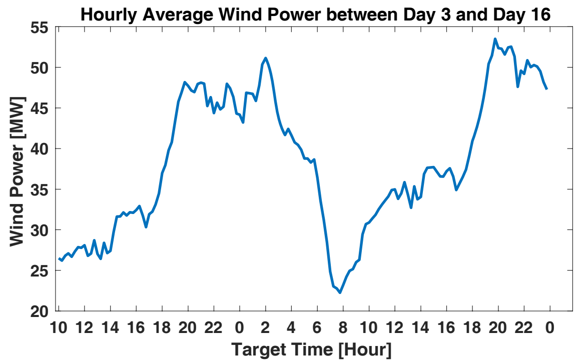

Fig. 7. The actual wind power between 2017/4/3 and 2017/4/16 is plotted

forecasting machines generally converges as the prediction horizon increases, if the average forecasting errors slightly decreases between 04:00 and 08:00, the forecasting performance temporally increases at that time.

5. Conclusion

In this study, we forecast the wind power after two hours up to 39 hours for 38 hours without the NWP. In order to utilize the historical weather information to increase the forecasting performance, we propose the new forecasting architecture based on the training data shift, where we assume that the future wind power depends on the far past weather information. We run the machine learning-based forecasting models on the new forecasting architecture. Furthermore, we also test the classical time series models, which forecast the future wind power recursively based on its own forecasts. In both processes, we use many forecasting techniques to increase the forecasting performance. Finally, we can conclude that the new forecasting architecture using the machine learning-based forecasting algorithm, which can also be called the regression-based forecasting algorithm, has a better performance than the classical time series models.

Acknowledgements

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. 20168530050080 and No. 20174030201660).

References

[1] Tao Hong, Pierre Pinson, Shu Fan, Hamidreza Zareipour, Alberto Troccoli, Rob J. Hyndman, "Probabilistic Energy Forecasting: Global Energy Forecasting Competition 2014 and Beyond," *International Journal of Forecasting*, vol. 32, no. 3, pp. 896-913, July 2016.

[2] Chen, Fei, and Jimy Dudhia. "Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and sensitivity," *Monthly Weather Review,* vol. 129, no. 4, pp. 569-585, Apr. 2001.

[3] Gneiting, Tilmann, and Adrian E. Raftery. "Weather forecasting with ensemble methods," *Science,* vol. 310, no. 5746, pp. 248-249, Oct. 2005.

[4] Athanasios Sfetsos, "A comparison of various forecasting techniques applied to mean hourly wind speed time series," *Renewable energy,* vol. 21, no. 1, pp. 23-35, Sep. 2000.

[5] Landry, Mark, Thomas P. Erlinger, David Patschke, and Craig Varrichio. "Probabilistic gradient boosting machines for GEFCom 2014 wind forecasting," *International Journal of Forecasting,* vol. 32, no. 3, pp. 1061-1066, Sep. 2016.

[6] 14th international Conference on the European Energy Market, 6-9, June 2017, Dresden, Germany, http://eem2017.com/program/forecast-competition

[7] Pierre Pinson, Christophe Chevallier, and George N. Kariniotakis. "Trading wind generation from short-term probabilistic forecasts of wind power," *IEEE Transactions on Power Systems,* vol. 22, no. 3, pp. 1148-1156, Aug. 2007.

[8] Friedman Jerome, Trevor Hastie, Robert Tibshirani, *The elements of statistical learning*, New York: Springer series in statistics, 2001.

[9] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, An introduction to statistical learning, New York: springer, 2013.

[10] Jerome H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics,* pp. 1189-1232, Oct. 2001.

[11] Box, George EP, Gwilym M. Jenkins, Gregory C. Reinsel, and Greta M. Ljung, *Time series analysis: forecasting and control,* John Wiley & Sons, 2015.

[12] Akaike, Hirotugu. "Factor analysis and AIC," *Psychometrika,* vol. 52, no. 3, pp. 317-332, Sep. 1987

[13] William Wu-Shyong Wei, *Time series analysis,* Addison-Wesley publication, 1994.

[14] System Identification Toolbox, MATLAB 2015b, MathWorks.

[15] Reinsel, Gregory C., *Elements of multivariate time series analysis,* Springer Science & Business Media, 2003.

[16] European Energy Market Competition Folder, https://cloudstore.zih.tu-dresden.de/index.php/s/DYt48551xFvUEqM, 2017.

Duehee Lee He received his B.S. in Electronic and Electrical Engineering in 2000 from Pohang University of Science and Technology, Pohang, Republic of Korea. He received his M.S. in the Electrical and Computer Engineering Department at The University of Texas at Austin, Austin, TX, in Dec 2009. He earned his PhD at the same University, in May 2015. His research interest is to integrate more wind power into the power system. He is currently an assistant professor in the KonKuk University, Seoul, Republic of Korea.

Yong-Gi Park He received B.S B.S., M.S., and Ph.D. degrees from Konkuk University, Korea, in 2005, 2009, and 2014 respectively. For 2005-2007, he worked in Samsung Electronics Corporation. For 2014-2017, he was with Research Center for Innovative Electricity Market Technology, Konkuk University, Korea. Currently, he is an assistant professor with Department of Electric and Electronic engineering, Youngsan University, Korea. His research interest includes optimal operation and planning of power system, and electricity market analysis.

Jong-Bae Park He received B.S., M.S., and Ph.D. degrees from Seoul National University in 1987, 1989, and 1998, respectively. For 1989-1998, he was with Korea Electric Power Corporation, and for 1998-2001 he was an Assistant Professor at Anyang University, Korea. For 2006-2008, he was a guest researcher of EPRI, USA. From 2001, he has been with Electrical Engineering Department at Konkuk University as Professor. His major research topics include power system operation, planning, economics, and markets.

Jae Hyung Roh He received the B.S. degree in Nuclear Engineering from Seoul National University, Korea, in 1993 and the M.S. degree in Electrical Engineering from Hongik University, Korea, in 2002. He received Ph.D. degree in Electrical engineering from Illinois Institute of Technology, Chicago, USA. For 1992-2001, he was with Korea Electric Power Corporation, and for 2001-2010, he was with Korea Power Exchange. Since 2010, he has been with Electrical Engineering Department at Konkuk University, Seoul, as an Associate Professor. His research interests include electricity market, smart grid and resource planning.