International Journal of Forecasting 30 (2014) 402–406

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Wind power forecasting using the *k*-nearest neighbors algorithm

E. Mangalova[a,*], E. Agafonov[b]

[a] *Siberian State Aerospace University, Russia*
[b] *Siberian Federal University, Russia*

ARTICLE INFO

Keywords:
Cross-validation
Data mining
Energy forecasting*
Forecasting competitions*
Feature selection
Nonparametric models
Regression tree

ABSTRACT

The paper deals with a modeling procedure which aims to predict the power output of wind farm electricity generators. The following modeling steps are proposed: factor selection, raw data pretreatment, model evaluation and optimization. Both heuristic and formal methods are combined to construct the model. The basic modeling approach here is the *k*-nearest neighbors method.

1. Introduction

The effective operation of wind power plants involves the optimization of their operating modes within the integrated energy system. In particular, it demands a prediction of the power output of each single plant. The problem statement and corresponding raw data were taken from the Global Energy Forecasting Competition 2012 (http://www.kaggle.com/c/GEF2012-wind-forecasting).

The data supplied include measurements of the following parameters from seven wind farms: weather forecasts with the meridional and zonal components of the wind, the angle of the wind, and the date and time of the measurement. The data cover a four-year period and consist of 18,757 measurements. Wind forecasts are available twice a day, with each one providing a prognosis for the next two days. Thus, we deal with multiple forecasts with different accuracies.

The data sets supplied are of two kinds: training and validation sets. The former serve as information for the predictive model development, the latter are provided for model quality estimation purposes.

The functioning of wind power plants tends to include irregular but frequent periods of downtime or reduced power delivery, probably as a result of routine maintenance, extraordinary weather, etc. The reasons for such periods are unknown, which leads to difficulties in the data analysis.

The solution to the problem of predicting power plants' electricity output includes the following steps: factor selection, raw data pretreatment, model evaluation and optimization. Both heuristic and formal methods are combined in order to construct the predictive model. The core modeling approach is the *k*-nearest neighbors method.

2. Model factors selection

The weather in the training data sets is represented by retrospective sequences of four forecasts. Starting with the selection of the model factors, we assume that the latest forecast is the most accurate, and therefore we simply remove older ones from the training data set.

The amount of electricity produced by a wind farm depends to a great extent on the air flow parameters. The power of air flow, in its turn, depends not only on the velocity of the flow, but also on the density of the air (Grogg, 2005). Unfortunately, we have no access to density-related parameters (such as temperature, humidity etc.) as such, though they may be expressed indirectly by some day/time information.

Finally, the following set of model factors is proposed for analysis: year, month, day of the month, day of the year, hour, zonal and meridional wind components, wind direction, and wind speed.

* Corresponding author.
E-mail address: e.s.mangalova@hotmail.com (E. Mangalova).

0169-2070/$ – see front matter © 2013 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijforecast.2013.07.008

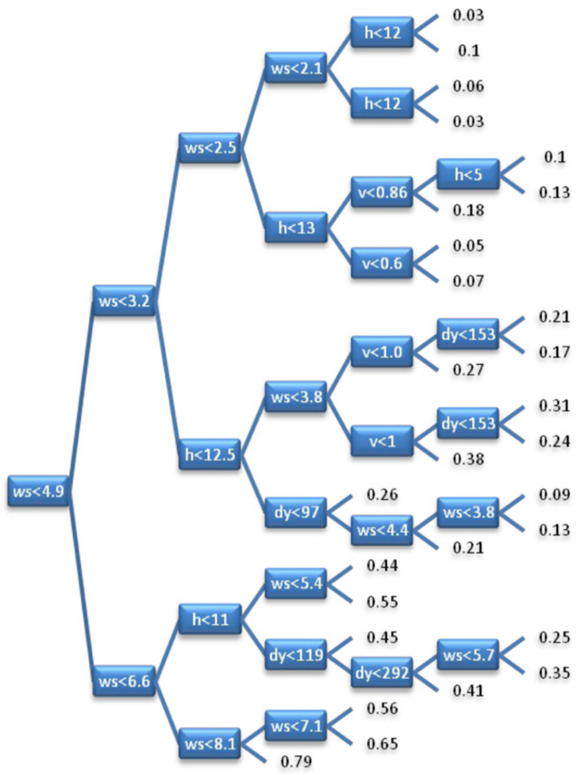

Fig. 1. The CaRT for wind farm 1: The lower alternative is for breaking the corresponding rule, the upper one is for satisfying the rule; *ws* is wind speed, *h* is hour, *dy* is day of the year, and v is the meridional wind component.

The CaRT (Classification and Regression Tree) approach (Breiman, Friedman, Olshen, & Stone, 1984) is applied for selecting the most significant factors from those listed above. The CaRT procedure splits the data set stepwise into subsets along those factors that allow for the best MSE improvement in the corresponding piecewise constant approximation. Thus, splitting along some factor indicates that the tree model has some dependence on this factor. This factor should be selected to appear in the final predictive model.

In addition to the above-mentioned factors, we also include the wind speed for neighbor wind farms. It is proposed that this be done later for individual wind farm models, while tuning them. Applying CaRT with the neighbor factors at the very beginning may lead to uncertainty in the selection of significant factors, due to the strong correlations between some of them, such as the weather forecast data (zonal and meridional wind components, wind direction, and wind speed).

As part of introducing the CaRT procedure, we also define a stopping rule. We require the number of data points in the leaves to be not less than 500. This stopping rule prevents the selection of factors which influence only small portions of the data set (less than about 5% of the training sample).

Fig. 1 depicts the CaRT for wind farm 1.

Table 1 contains plus signs for the factors and wind farms that have been split during the regression tree construction, and have therefore proved their significance.

Table 1
Significant factors for wind farms, found using CaRT.

	1	2	3	4	5	6	7
Zonal wind component		+	+	+		+	+
Meridional wind component	+	+	+	+	+	+	+
Wind direction		+			+		+
Wind speed	+	+	+	+	+	+	+
Year					+		
Month							
Day of the month							
Hour	+	+	+	+	+	+	+
Day of the year	+	+			+	+	+

Any factors which were significant for five or more wind farms were chosen for inclusion in the final predictive model.

Let us denote the basic set of significant factors as follows: x^1 is the zonal wind component, x^2 is the meridional wind component, x^3 is the wind speed, x^4 is the hour, and x^5 is the day of the year. The wind speeds for neighboring wind farms become the complimentary factors for individual wind farm models. We add neighboring wind speeds into the model sequentially: the wind speed x^6 should provide the greatest improvement in model accuracy, and x^7 is chosen so as to proceed with the improvement.

In addition, let y denote the normalized wind power measurements and n the sample size.

3. Raw data pretreatment

Weather forecasts contain the most significant information for the model process, but they are stochastic. We propose the following procedures for weather forecast pretreatments:

(a) Algorithm of the simple moving average:

Step 1:

Assign

$$\widehat{x}^j = x^j = \left(x_1^j, x_2^j, \ldots, x_n^j\right), \quad j = 1, 2, 3, 6, 7. \tag{1}$$

Step 2:

$$x_p^j = \frac{\sum_{i=-t_1}^{t_2} \widehat{x}_{p+i}^j}{t_1 + t_2 + 1}, \quad j = 1, 2, 3, 6, 7,$$
$$p = t_1 + 1, t_1 + 2, \ldots, n - t_2, \tag{2}$$

where $t_1, t_2 \geq 0$, $[t_1, t_2]$ is the smoothing interval.

(b) Algorithm of the weighted moving average:

Step 1:

Assign

$$\widehat{x}^j = x^j = \left(x_1^j, x_2^j, \ldots, x_n^j\right), \quad j = 1, 2, 3, 6, 7. \tag{3}$$

Step 2:

$$x_p^j = \frac{\sum_{i=-t_1}^{t_2} \omega_{t_1+i+1} \widehat{x}_{p+i}^j}{\sum_{i=-t_1}^{t_2} \omega_{t_1+i+1}}, \quad j = 1, 2, 3, 6, 7,$$
$$p = t_1 + 1, t_1 + 2, \ldots, n - t_2, \tag{4}$$

where ω are weights.

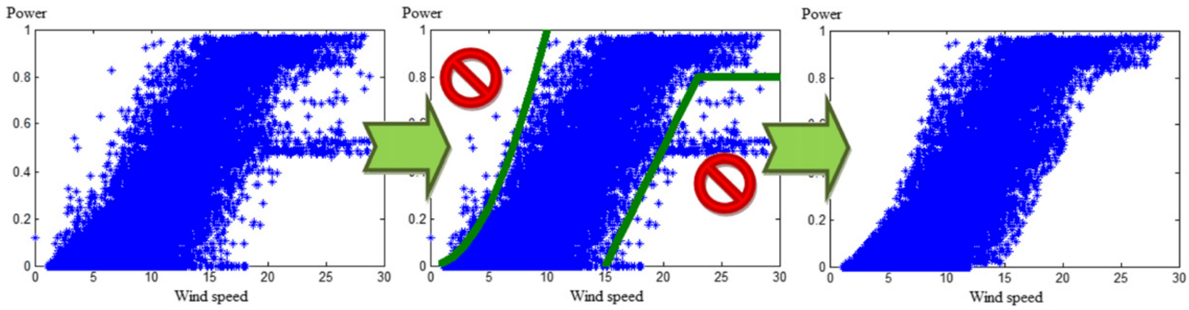

Fig. 2. Censoring procedure.

(c) Modified moving average:

The modified moving average is calculated consistently for each p:

$$x_p^j = \frac{\sum_{i=-t_1}^{t_2} x_{p+i}^j}{t_1 + t_2 + 1}, \quad j = 1, 2, 3, 6, 7,$$
$$p = t_1 + 1, t_1 + 2, \ldots, n - t_2. \tag{5}$$

Smoothing algorithm (c) demonstrates the best effectiveness according to both the forecast errors and the computation speed, because it contains only two tunable parameters and does not demand an extra memory allocation.

After going through the smoothing procedure, one revises the data set by removing any abnormal data points.

Electric generators should possess a monotonically rising "power vs. angular rotor speed" dependence, but some fragments of the training data set fail to conform to this. Hence, abnormal measurements of the corresponding parameters are suspected. Another possibility is that these data points may correspond to times of substandard operation of the power plant.

Two cases are reviewed:

(1) normalized wind power is high with slow wind, and
(2) normalized wind power is low with strong wind.

Taking outliers into consideration inevitably influences the quality of the corresponding model. Censoring data is a common procedure for improving the resulting model. Introduce the following censoring rule: if $y_i > a_1 \left(x_i^3\right)^2$ or $y_i < a_2 x_i^3 \cap y_i < a_3$ $(i = 1, 2, \ldots, n)$, then the ith element should be removed from the learning data set. Fig. 2 depicts this censoring procedure.

The parameters α_1, α_2 and α_3 are found for each wind farm individually.

4. Model evaluation and optimization

For the prediction of the normalized wind power, we propose the k-nearest neighbors algorithm (Hardle, 1990). This choice is made for the following reasons:

(a) *Interpretability of the model*. The results of the predictive algorithm using the k-nearest neighbors approach are based upon the occasions in the past that are closest to the current state (according to a given distance metric). Prediction is fulfilled by a simple averaging of the output values of the k nearest neighbors, or by some weighted averaging. The k-nearest neighbors algorithm allows its results to be interpreted by experts.

(b) *Cyclic factors treatment*. The factors used include some cyclic ones (year, month etc.). The k-nearest neighbors algorithm can be tuned to work with them (unlike tree methods, for example, which are not able to deal with cyclic factors).

(c) *No multiple learning* is needed with the k-nearest neighbors algorithm when new portions of data are introduced. In this case, when adding samples, we expand the search instances without a need to recalculate the model. Removing old data is also done without repeated learning.

The following distance metrics are proposed for the selection of the neighbors.

(1) Single factor metric:

$$d^j\left(x_p^j, x_q^j\right) = \left|x_p^j - x_q^j\right|, \quad j = 1, 2, 3, 6, 7,$$
$$p = 1, 2, \ldots, n, \; q = 1, 2, \ldots, n. \tag{6}$$

(2) Cyclic single factor metric:

$$d^4\left(x_p^4, x_q^4\right) = \min\left(\left|x_p^4 - x_q^4\right|, 24 - \left|x_p^4 - x_q^4\right|\right),$$
$$p = 1, 2, \ldots, n, \; q = 1, 2, \ldots, n, \tag{7}$$

$$d^5\left(x_p^5, x_q^5\right) = \min\left(\left|x_p^5 - x_q^5\right|, 365 - \left|x_p^5 - x_q^5\right|\right),$$
$$p = 1, 2, \ldots, n, \; q = 1, 2, \ldots, n. \tag{8}$$

(3) Multiple factor metric:

$$D\left(\bar{x}_p, \bar{x}_q\right) = \sum_{j=1}^{7} w_j d^j\left(x_p^j, x_q^j\right),$$
$$p = 1, 2, \ldots, n, \; q = 1, 2, \ldots, n, \tag{9}$$

where w_j are the weights of the corresponding factors.

Factors 1 and 2 (the zonal and meridional wind components, respectively) are measured on the same scale, and we therefore let them have the same weights, $w_1 = w_2$.

Finally, we arrived at the following predictive model:

$$\hat{y}\left(\bar{x}\right) = \frac{\sum_{q=1}^{n_{\tau-1}} \varphi\left(\bar{x}, \bar{x}_q\right) y_q}{\sum_{q=1}^{n_{\tau-1}} \varphi\left(\bar{x}, \bar{x}_q\right)}, \tag{10}$$

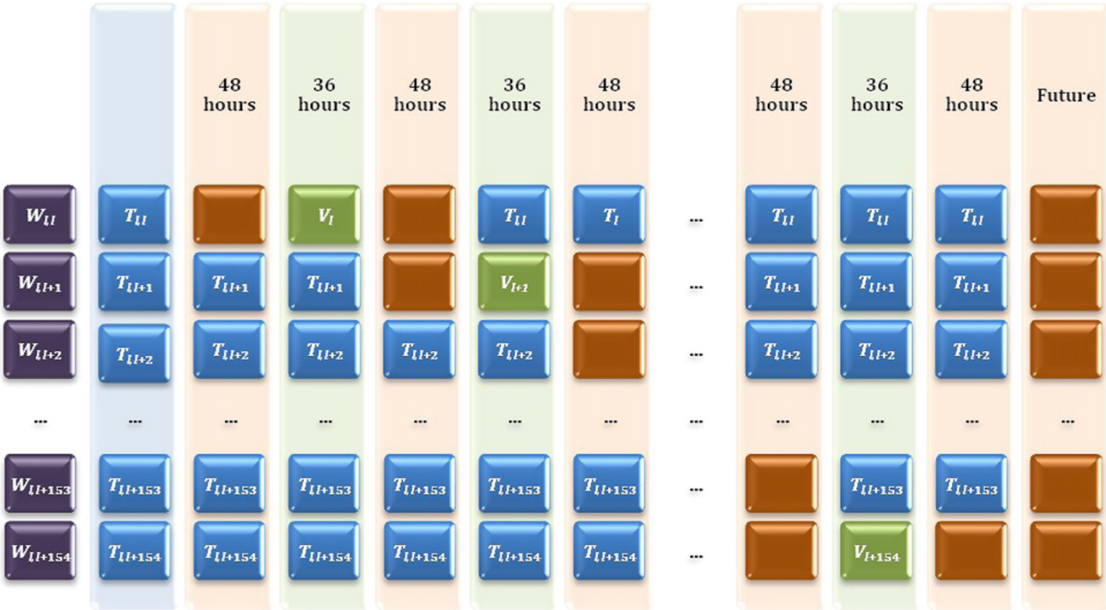

Fig. 3. The construction scheme from Eq. (15) and Eq. (16). The observation sets marked red are included in neither the validation set (V) nor the training set (T). (For the interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where $n_{\tau-1}$ is the number of the last data point received,

$$\varphi\left(\bar{x}, \bar{x}_q\right) = \begin{cases} D\left(\bar{x}, k\right) - D\left(\bar{x}, \bar{x}_q\right) + 0.1, \\ \quad D\left(\bar{x}, \bar{x}_q\right) \le D\left(\bar{x}, k\right), \\ 0, \quad D\left(\bar{x}, \bar{x}_q\right) > D\left(\bar{x}, k\right), \end{cases} \tag{11}$$

and $D\left(\bar{x}, k\right)$ is the distance between $\bar{x}$ and the kth nearest neighbor.

Steady weather leads to the appearance of close instances in the data set. This effect means that previously received data instances are neighbors of the one to be predicted, which leads to the problem of overfitting. For short-term prediction, the model is very accurate, but it fails for long-term prognosis. In order to solve the problem by validating the predictive model in Eq. (10), we apply modified q-fold cross-validation. This proposed cross-validation should avoid the issue of taking previous instances from the neighboring set.

The data nearest to the validation set are therefore excluded from the training set. Thus, we get rid of "*good in advance*" neighbors to prevent overfitting.

In order to optimize the model in Eq. (10), we propose the following criterion:

$$W\left(\bar{w}, \bar{t}, k\right) = \sum_{l} \sum_{i \in V_l} \left(y_i - \hat{y}\left(\bar{x}_i, T_l\right)\right)^2, \tag{12}$$

where $V_l = \left(\left(\bar{x}_{z_1^l}, y_{z_1^l}\right), \left(\bar{x}_{z_2^l}, y_{z_2^l}\right), \ldots, \left(\bar{x}_{z_v^l}, y_{z_v^l}\right)\right)$ are validation sets, $l = 1, 2, \ldots, L$, $z_i^l \in N$, $z_i^l \le n$, $i = 1, 2, \ldots, n$, and the k nearest neighbors of $\bar{x}_i$ are taken from the set $T_l = \left(\left(\bar{x}_f, y_f\right) : \forall\left(\bar{x}_{z_i^l}, y_{z_i^l}\right) \in V_l \,|z - f| > 48\right)$.

For any argument set of goal functions, we choose t and the number of neighbors k by implementing an exhaustive search (where t_1 and t_2 are chosen from the range [0, 3], and k from the range [1, 250]).

Optimization with respect to parameters w is achieved by using the modified coordinate-wise descent. We begin with a randomly-chosen point $\bar{w} = (w_1, w_2, w_3, w_4, w_5, 0, 0)$, where $w_j \in (0, 1)$, $j = 1, 2, \ldots, 5$. Changing the first five nonzero weights sequentially, one moves towards the minimum of the criterion in Eq. (12). Having reached the local minimum $W^* = W\left(\bar{w}^*, \bar{t}, k\right)$, one varies the sixth (previously zero) weight in the model by adding the wind speeds for the neighbor wind farms one by one, and continue moving towards the minimum. From the six new models, we choose the best one according to the criterion in Eq. (12) $\left(W^{**} = W\left(\bar{w}^{**}, \bar{t}, k\right)\right)$. If W^{**} for the best model is better than W^*, we keep the model for the further improvement. We then add the seventh weight for the other five neighboring farms' wind speeds, providing optimization. From these five new models, we again choose the best one according to the criterion in Eq. (12) $\left(W^{***} = W\left(\bar{w}^{***}, \bar{t}, k\right)\right)$, if W^{***} for the best model is better than W^{**}.

While tuning the parameters, we optimize the criterion in Eq. (12) with different validation sets. We generate the validation and training sets as follows:

$$V_l = \left(\left(\bar{x}_\lambda, y_\lambda\right), \left(\bar{x}_{\lambda+1}, y_{\lambda+1}\right), \ldots, \left(\bar{x}_{\lambda+35}, y_{\lambda+35}\right)\right), \quad l = 1, 2, \ldots, 310, \tag{13}$$

where $\lambda = 122 + 84\,(l - 1)$,

$$T_{l,l+b-1} = \left(\left(\bar{x}_f, y_f\right) : \forall(\bar{x}_z, y_z) \in V_l |z - f| > 48 \cap f \in [1122 + 84(l + 154)]\right), \quad l = 1, 2, \ldots, 155, \; b = 1, 2, \ldots, 155. \tag{14}$$

Grouping the criteria for 155 sequential validation sets, we generate the new criteria (Fig. 3):

$$\tilde{W}_l = \sum_{b=1}^{155} W_{l,l+b-1}\left(\bar{w}, \bar{t}, k\right), \quad l = 1, 2, \ldots, 156, \tag{15}$$

Table 2
Optimal parameters.

Wind farm #	k	w^1, w^2	w^3	w^4	w^5	w^6		w^7	
						Wind farm #	w^6	Wind farm #	w^7
1	81	0.55	2.5	0.35	0.035	4	1.1	7	0.8
2	151	0.65	3	0.65	0.025	1	0.73	–	–
3	96	0.45	1.45	0.3	0.025	7	0.1	–	–
4	112	0.3	2.5	0.25	0.03	7	1.3	1	0.4
5	67	0.6	2.2	0.25	0.016	1	0.18	–	–
6	138	0.3	2.1	0.25	0.031	7	2.25	–	–
7	110	0.27	2	0.12	0.016	4	0.4	–	–

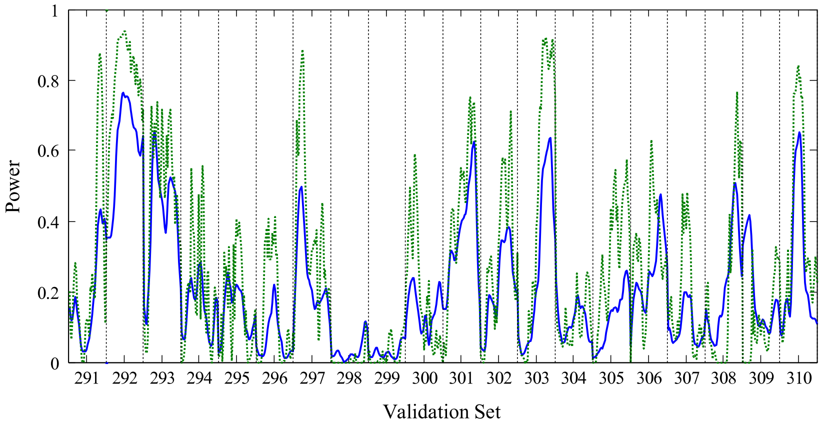

Fig. 4. Normalized wind power forecasting results for the validation sets 291–310: the validation set (dashed line) and forecast (solid line).

where

$$W_{l,l+b-1}\left(\bar{w}, \bar{t}, k\right) = \sum_{i \in V_l} \left(y_i - \hat{y}\left(\hat{x}_i, T_{l.l+b-1}\right)\right)^2,$$

$$l = 1, 2, \ldots, 156, \; b = 1, 2, \ldots, 155. \tag{16}$$

The global minima of Eq. (15) for different validation sets do not coincide, but are always in the same neighborhood. The optimal criteria values always stay within the range of 0.07%, which proves the stationary nature of the process under investigation.

Table 2 contains the optimal parameters of the model built according to the last criterion from the set in Eq. (15). The parameter $\bar{t}$ is the same for all of the wind farms: $\bar{t} = (2, 1)$.

In order to improve the model accuracy, we smooth the model along the time axis, implementing the simple moving average:

$$\tilde{y}\left(\bar{x}_p\right) = \frac{\sum_{i=-2}^{2} \hat{y}\left(\bar{x}_{p+i}\right)}{5}. \tag{17}$$

If the values of the normalized wind power for time instances $p-2$ and $p-1$ (y_{p-2} and y_{p-1}) are known, we propose to use them instead of $\hat{y}\left(\bar{x}_{p-2}\right)$ and $\hat{y}\left(\bar{x}_{p-1}\right)$ in Eq. (17). The simple moving average reduces the modelling errors which are influenced by weather forecast time shifts.

The final model in Eq. (17) is able to accumulate information by an increase of the neighbors' space. Fig. 4 shows normalized wind power forecasting results for the validation sets 291–310 (wind farm #1).

The model in Eq. (17) has been verified using a test sample. The accuracy (RMSE) of the power prediction of wind farms has reached a level of 0.1472 in the case of full data availability ($n_{\tau-1} = n$ in Eq. (10)) and 0.1502 in the case of sequential data accumulation (Eq. (10)).

5. Conclusion

This work is devoted to the problem of the prediction of wind power plants' power output. Nonparametric k-nearest neighbors has been applied as the main modeling approach. Based on the RMSE criterion, the predictive model achieved second place in the Global Energy Forecasting Competition 2012.

The resulting model allows the prediction of individual wind power plants' outputs. An accurate prognosis of the electric power delivered, together with information on power consumption, provides a great tool for the minimization of reserve power usage. It is important to note that today's reserve power facilities consume organic fuel, and that each startup of the reserve plants is both time-consuming and expensive. The accurate prediction of wind power plant operation could prevent unnecessary usage of the backup power plants, or, when necessary, prepare them in advance.

References

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth Inc..
Grogg, K. (2005). *Harvesting the wind: the physics of wind turbines*. Carleton College Senior Comprehensive Papers in Physics and Astronomy.
Hardle, W. (1990). *Applied nonparametric regression*. Cambridge University Press.