Available online at www.sciencedirect.com
ScienceDirect
Energy Reports 7 (2021) 332–338
www.elsevier.com/locate/egyr

The 4th International Conference on Electrical Engineering and Green Energy CEEGE 2021, 10–13 June, Munich, Germany

Examination of turbulence impacts on ultra-short-term wind power and speed forecasts with machine learning

Hao Chen*, Yngve Birkelund, Fuqing Yuan

Department of Technology and Safety, UiT The Arctic University of Norway, Tromsø 9019, Norway

Received 20 July 2021; accepted 5 August 2021

Abstract

Wind turbines' economic and secure operation can be optimized through accurate ultra-short-term wind power and speed forecasts. Turbulence, considered as a local short-term physical wind phenomenon, affects wind power generation. This paper investigates the use of turbulence intensity for ultra-short-term predictions of wind power and speed with a wind farm in the Arctic, including and excluding wind turbulence, within three hours by employing several different machine learning algorithms. A rigorous and detailed statistical comparison of the predictions is conducted. The results show that the algorithms achieve reasonably accurate predictions, but turbulence intensity does not statistically contribute to wind power or speed forecasts. This observation illustrates the uncertainty of turbulence in wind power generation. Besides, differences between the types of algorithms for ultra-short-term wind forecasts are also statistically insignificant, demonstrating the unique stochasticity and complexity of wind speed and power.

Peer-review under responsibility of the scientific committee of the 4th International Conference on Electrical Engineering and Green Energy, CEEGE, 2021.

Keywords: Machine learning; Statistical comparison; Turbulence; Wind energy; Wind forecast

1. Introduction

Establishing accurate wind power prediction models is of great significance to the power grid's safe and stable operation and economic operation [1]. Moreover, from the perspective of power generation companies, accurate and reliable prediction of wind energy in the short term is of great importance for the efficient operation of wind farms [2]. It can also prompt them to participate in electricity market competition [3], reduce economic losses caused by electricity supply uncertainties, and make reasonable wind farms' practical maintenance plans. Wind power forecasting can describe wind characteristics and power in the next minutes, hours, days, or even weeks based on wind farms or meteorological data. This paper focuses on ultra-short-term forecasts (a few seconds to 4 h) used for turbine control and load tracking [4].

* Corresponding author.
E-mail address: hao.chen@uit.no (H. Chen).

https://doi.org/10.1016/j.egyr.2021.08.040

Peer-review under responsibility of the scientific committee of the 4th International Conference on Electrical Engineering and Green Energy, CEEGE, 2021.

The research for ultra-short-term can be considered forecasting of a time series and thus ignores the meteorological factors. Gangui Y et al. (2012) [5] took ultra-short-term wind power production as a multiple chaotic time series problem. They used a validation of their solution using a real wind farm in northeast China using forecasting times of 15 min, 30 min, and 1 h. Zhang Z Z et al. (2011) [6] proposed an improved GM (Grey Model) to forecast ultra-term wind speed. It used the relationship between wind speed and wind power to make a prediction. Utilizations of different learning algorithms for forecasting wind are also prevalent. Shi K et al. (2018) [7] also demonstrated the enhanced accuracy, efficiency, and robustness of improved random forests for short-term wind power forecasting, which has better performance than the backpropagation neural network, Bayesian network, and support vector machine. Lee J et al. (2020) [8] compared ensemble learning-based models in the wind power prediction on ten minutes of data from actual wind turbines located in France and Turkey. It showed that the ensemble methods could predict wind power production with high accuracy than the standalone machine learning models. These investigations are normally algorithm-oriented and the benchmark algorithms for comparing the proposed algorithms are often of the same type, without cross-algorithm comparisons.

There are a few studies about turbulence in wind power forecasting. Nielson J et al. (2020) [9] set up an artificial neural network with wind speed, density, Richardson number, turbulence intensity, and wind shear as input parameters to improve wind turbine power prediction. Li F et al. (2019) [10] conducted a multistep wind speed prediction using turbulence into the hybrid deep neural networks on multiple prediction intervals from 10 min to 12 h and finding the higher resolution turbulence intensity incorporated in good wind prediction. However, these studies typically claim that models that consider turbulence make more accurate predictions, but their results are not tested statistically.

This paper uses a rigorous statistical approach to test whether turbulence has a notable role in wind power and speed forecasts and compares the performance of different types of machine learning predictive algorithms.

2. Wind turbulence and data preparation

Wind energy is a form of conversion of solar energy: the solar radiation energy received by the Earth is converted into wind energy by temperature gradients in the air [11]. Wind power generation is the process of converting wind energy into electrical energy. As a local wind phenomenon, turbulence has a significant impact on wind turbine electricity generation in wind park operations. Due to the uneven terrain or air density difference, the airflow will generate turbulence when flowing. On similar wind speed conditions, the higher the turbulence intensity, the higher the impact of wind farm output power [12]. At low wind speeds, turbulence increases the electrical power production of the turbine. However, when the wind speed approaches the turbine's furling speed, turbulence reduces energy production [13]. In statistics, the standard deviation measures the amount of variation or dispersion of a set of values. Turbulence is an extremely complex fluid phenomenon with intense randomness that is difficult to describe precisely. Turbulence intensity is one of the main characteristics quantity of wind speed fluctuations. It is defined as dividing the standard deviation of wind speed by the mean wind speed in a short time interval [14]. In this research, we define turbulence intensity I_i within ten minutes intervals i as: $I_i = S_i/SP_i$, where SP_i is wind speed, and S_i is its standard deviation of the previous ten minutes.

The meteorological wind data measurements are from a wind park, with an installed capacity of 54 MW with 18 Vestas V90 3.0 MW turbines, flat hills and towards a fjord, and an average altitude of 95 m. It is a whole year data from 0:00 1st January 2017 to 23:50 31st December with ten minutes temporal resolution. The size of the data sample is 52,560. Since the ranges of variables of the data set are quite different, it is necessary to rescale the raw data into new data with a similar scale of each variable. There are standard data rescaling methods, namely normalization, and stabilization. In this research, we choose stabilization, by subtracting the overall average from the original data and dividing the difference by the standard deviation. Consequently, it rescales original data to a new data set with a mean of zero and a standard deviation of one.

3. Methodology

This section presents four well-performing, representative machine learning algorithms for wind power and speed forecasts and metrics to evaluate their predictive performance. Besides, statistical methods for comparing their results are also described.

Linear Regression (LR): Linear regression algorithm is a basic supervised machine learning algorithm due to its relatively simple and well-known characteristics. It uses a least-squares function named linear regression equation to model the relationship between independent and dependent variables. This function is a linear combination of one or more model parameters called regression coefficients [15].

Back Propagation Neural Network (BPNN): The neural network is a bionic machine learning algorithm inspired by the biological neural networks that constitute animal brains. Besides, it enables these models to solve prediction problems with nonlinear structures. It is proven its edge in wind prediction problems [16]. For BPNN, a typically three-layered structure consists of input, hidden, and output layers, and the loss function gradients are computed and backpropagated. In this study, the BPNN comprises 20 nodes of the hidden layer and one node output layer.

Reduced-Error Pruning TREE (REPTREE): The decision tree is a popular predictive machine learning algorithm because of its understandability and simplicity. A decision tree generated by the algorithm is typically large for a big data set, and each variable has been considered in detail. It may raise the problem of overfitting. REPTREE is a practical decision tree pruning method that sets a new validation to correct the tree to overcome the overfitting problem [17]. It traverses all the subtrees sequentially from bottom to top. A new, relatively simplified decision tree is created for each subtree of a non-leaf node replaced with a leaf node. As a result, the terminated pruning algorithm typically offers a more superficial and more generalized decision tree.

Random Forest (RF): Bagging is a unique algorithm of the model averaging approach to reduce the prediction variances by using repetitions of creating multiple sets of original data to train the machine learning model. Random Forest (RF), proposed by Ho in 1995 [18], is an efficient ensemble machine learning. RF is based on the construction of many basis learner. Each tree is trained by using a bootstrap sample extracted from the whole training set. The forest of regressions produces an ensemble value. The final regression value can be determined in kinds of averages [19].

The ultra-short-term wind forecasting employs a predictive variable autoregression strategy in conjunction with other variables, like turbulence intensity, to complement the forecasting analysis. This strategy allows the adequate exploitation of predictive variables' time-series information and absorbs information from other variables to improve the forecast model. The general forecast as step i+n is described as:

$$\hat{y}_{i+n} = f\left(y_{i-1}, \ldots, y_{i-6}; \vartheta_{i-1}, \ldots, \vartheta_{i-6}\right) + \varepsilon_n \tag{1}$$

where $\hat{y}_{i+n}$ is n time steps ahead predictive wind variable, ϑ represents assistant variables that may offer additional information in predictive models, ε_n is the error of the model. Given the data's temporal resolution and the short-term property of turbulence, the furthest previous data are set to one hour before the current time, six-time steps before. Besides, the maximum forecast time is chosen as three hours, which is eighteen-time steps ahead.

There are two metrics in evaluating forecast performance with different machine learning algorithms. Namely, Root Mean Square Error (RMSE) and Mean Directional Accuracy (MDA). The first is error magnitude metrics, and the second is an error direction index, which is used in econometrics but rarely in energy science. Besides, $\mathbf{1}_{sgn(.)}$ is the indicator function in Eq. (3).

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(y(t) - \hat{y}(t)\right)^2} \tag{2}$$

$$MDA = \frac{1}{n}\sum_{t=1}^{n}\mathbf{1}_{sgn\left(y(t)-\hat{y}(t)\right)} \tag{3}$$

Three statistical methods are used to test whether there are statistically significant differences between results in different this study. Viz. Paired T-test, analysis of variance (ANOVA), and Tukey method for confidence intervals (CIs) between means of two populations [20]. The first is for paired comparisons, and the other two are for multiple comparisons. For the two tests, their hypotheses are similar. H_0: The means of these populations are equivalent; H_a: At least one does not equal the other. Their test statistics are as below:

$$T = \frac{\overline{Y}_1 - \overline{Y}_2}{S_{(\overline{Y}_1 - \overline{Y}_2)}} \sim t_{n_1+n_2-2} \tag{4}$$

$$F = \frac{Variance\ between\ groups\ differences}{Variance\ within\ groups\ differences} \sim F_{k,n-k} \tag{5}$$

The Tukey method for CIs is expressed as:

$$\left(\overline{Y}_{1} - \overline{Y}_{2\cdot}\right) \pm \frac{q_{k,n-k,1-\alpha}}{\sqrt{2}} \cdot \sqrt{MSE} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \tag{6}$$

where S is the standard deviation, t and q are t and Gaussian q-distributions, k is the number of populations and n is the total size of all populations, and *MSE* is the mean square error within groups.

4. Experimental results and discussions

To test whether turbulence makes a significant difference in ultra-short-term wind prediction. We perform multistep predictions of wind power and wind speed itself separately with the above algorithms. The procedure is illustrated in Fig. 1. Given the relatively large sample size, the testing set is configured as one-tenth of the total sample. This paper is concerned with ultra-short-term forecasting; half an hour, one, and three hours are selected as the maximum prediction timesteps, and results are tallied. The results are compared with the statistical method mentioned previously. (Note: The following p-values are less than the -6th power of 10 shortened to 0.)

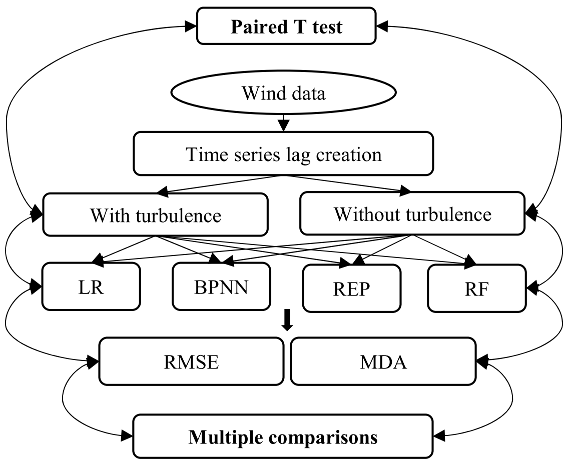

Fig. 1. Procedure for wind forecasts and statistical tests.

4.1. Wind power forecast

Four machine learning algorithms are applied for multistep predictions of wind power. The first of these prediction models include wind speed turbulence intensity, and the second (marked with *) does not. Table 1 shows RMSE and MDA of three-time steps wind power forecasts with LR, BPNN, REPTREE, and RF algorithms, including and excluding turbulence of wind speed.

Table 1. The performance of three steps ahead wind power forecasts with machine learning algorithms.

Metrics	Step1	* Step1	Step2	* Step2	Step3	* Step3
LR RMSE	0.2331	0.2331	0.3474	0.3473	0.4028	0.4028
LR MDA	57.7231	58.0854	48.0641	48.0832	48.2068	48.2259
BPNN RMSE	0.2307	0.232	0.3447	0.3447	0.4013	0.4009
BPNN MDA	57.4371	58.0854	48.0259	48.3883	49.2751	48.512
REPTREE RMSE	0.2434	0.2429	0.3575	0.3574	0.4166	0.4165
REPTREE MDA	39.2449	39.4928	30.9746	30.9556	30.5609	30.5799
RF RMSE	0.2496	0.252	0.3713	0.3704	0.4327	0.4343
RF MDA	55.8924	55.3013	48.4646	47.7017	48.4357	48.016

It is shown that as the forecasting step increases, the RMSE of two cases of all algorithms raises, and the metric increases slower for each step. There is no clear trend in the variation of MDA. From the first inspections of these results, forecast models with and without turbulence do not perform differently with the same algorithms. The results for the four algorithms are quite similar. To rigorously verify whether wind speed turbulence has a significant effect on wind power prediction, paired T-tests are conducted for the results of models built on the same forecasting algorithm, respectively. The *p*-values are shown in Table 2. It is seen that for three and six-time steps, the *p*-values are higher than 0.05 for almost all tests, indicating there is statistical evidence that the inclusion and exclusion of turbulence density do not have significant impacts on ultra-short-term wind power forecasts in these cases. It is notable that when the forecast time is extended to three hours, the models' performance with and without turbulence appears to some differences. Therefore, it cannot be inferred whether counting the turbulence term improves the model accuracy or adds noise to the power prediction.

Table 2. The *p*-values of paired T-tests for time steps (metric *plus* 'steps') ahead wind power forecasts.

Metrics (no. means timesteps)	LR vs. LR*	BPNN vs. BPNN*	REP vs. REP*	RF vs. RF*
RMSE 3	0.423	0.618	0.222	0.408
MDA 3	0.364	0.866	0.426	0.027
RMSE 6	0.025	0.713	0.315	0.051
MDA 6	0.371	0.602	0.792	0.075
RMSE 18	0	0.584	0	0.223
MDA 18	0.62	0.395	0	0.016

4.2. Wind speed forecast

Analogously to wind power prediction, models containing and not containing turbulence are constructed, and multistep wind speed predictions are performed. The metrics for forecasts are displayed in Table 3. These metrics temporal alterations for wind speed forecasts are similar to their counterparts in power forecasts cases.

Table 3. The performance of three steps ahead wind speed forecasts with machine learning algorithms.

Metrics	Step1	* Step1	Step2	* Step2	Step3	* Step3
LR RMSE	0.228	0.2282	0.3073	0.3079	0.3482	0.3492
LR MDA	46.9458	46.8696	43.2242	43.1861	44.3556	44.2604
BPNN RMSE	0.2247	0.228	0.3045	0.3067	0.3474	0.347
BPNN MDA	47.6689	47.0219	43.3003	43.5097	45.1171	44.9267
REPTREE RMSE	0.2382	0.2381	0.3222	0.3219	0.3642	0.3647
REPTREE MDA	36.3654	35.6232	31.7472	31.8234	33.676	33.3524
RF RMSE	0.2316	0.2373	0.3114	0.32	0.3561	0.3611
RF MDA	46.7555	46.2226	43.6049	44.4233	45.3455	46.126

Likewise, the paired T-tests are made to check the turbulence function in multistep speed predictions in Table 4. These tests for three and six-time steps wind speed forecasts also reject the null hypothesis and verify turbulence intensity's ineffectiveness. However, turbulence statistically changes the overall performance of predictive models for 3 h (18 time steps) ahead of forecasts.

Table 4. The *p*-values of paired T-tests for time steps (metric *plus* 'steps') ahead wind speed forecasts.

Metrics (no. means timesteps)	LR vs. LR*	BPNN vs. BPNN*	REP vs. REP*	RF vs. RF*
RMSE 3	0.122	0.261	0.902	0.028
MDA 3	0.053	0.487	0.297	0.508
RMSE 6	0.008	0.703	0.153	0
MDA 6	0.111	0.177	0.367	0.256
RMSE 18	0	0.001	0	0
MDA 18	0.851	0.453	0.130	0.368

4.3. Multiple comparisons between forecast algorithms

To scientifically investigate the differences between machine learning algorithms for wind power and wind speed forecasts, ANONA is carried out among the various metric, corresponding to eighteen steps predications with turbulence. These algorithms and results are presented in Table 5. It turns out that there is no substantial difference in the performance of these forecast algorithms, as a group, for both wind power and speed predictions regarding RMSE since their p-values are considerably larger than 0.05. Among them, the smaller p-values corresponding to forecasting wind power forecasts indicate that differences in forecasting wind power with these algorithms are more insignificant compared to wind speed.

Table 5. The multiple comparisons of eighteen steps ahead wind power and speed forecasts with turbulence.

Statistics	Power RMSE	Speed RMSE	Power MAD	Speed MAD
F	0.863	0.245	395.881	687.393
p-value	0.464	0.865	0	0

Moreover, multiple pair comparisons of metrics with Tukey methods also prove that no difference in RMSE is found between these prediction algorithms in forecasting wind power and speed since confidence intervals for their differences all contain zero. In particular, from Table 6, the REPTREE algorithm statistically shows lower MDAs in both forecasts, suggesting that its prediction error distribution is more symmetrically distributed than other algorithms, with zero centered.

Table 6. The bounds with 95% CIs for paired comparisons of MDA for wind power and speed forecasts algorithms.

Bounds	LR vs. BPNN	LR vs. REP	LR vs. RF	BPNN vs. REP	BPNN vs. RF	REP vs. RF
Power lower	−1.5265	16.4945	−1.0885	16.3425	−1.2405	−19.2615
Power upper	1.8304	19.8515	2.2684	19.6995	2.1165	−15.9046
Speed lower	−1.0981	11.9981	−1.1182	12.1695	−0.9469	−14.0431
Speed upper	0.7554	13.8516	0.7352	14.0229	0.9066	−12.1896

5. Conclusion

Ultra-short-term wind forecasting is essential for optimal control and operational efficiency of wind turbines. Turbulence in the wind has implications on wind power generation. In the present study, we focus on various machine learning autoregressive approaches to realize forecasts for wind power and speed for a wind farm inside the Norwegian Arctic regions. The effects of turbulence terms in modeling and different algorithms are compared.

The performances of different machine learning algorithms in predicting ultra-short-term wind power and speed are satisfactory but not significantly different in general. Their error distributions are different to some extent. This phenomenon may be interpreted as an absence of apparent variations of variables in the ultra-short-term. These variations are quite stochastic, resulting in the time series resembling a random walk in a short period so that prediction algorithms hardly capture their patterns. According to the statistical analysis, no clear statistical evidence exists that wind speed turbulence intensities affect the ultra-short-term wind power and speed forecasts. The main reason is that in ultra-short-term forecasts, the predictor variable's previous data are the most dominant factor affecting their predictive values, and other variables serve only as supplementary information. It suggests that it might be ill-advised to directly employ turbulence intensity into the forecast model, given that it is a subsidiary factor and increases computational burdens.

Since the wind farm understudy has a complex topography, there may be turbulence interactions, both natural and generated by the wind turbines. As a whole wind farm, these turbulent currents could cancel each other out. It is advantageous to conduct the examination of turbulence effect for a single wind turbine. Even though the effect of wind speed turbulence intensity is not significant in our case, it is still detected that it has a greater impact on ultra-short-term wind speed prediction than power, which indicates that there are interactions between weather factors. It also implies that if wind speed, turbulence, and other weather factors impacting wind power generation are taken into account in an appropriate methodology, wind power forecasts accuracy may be improved. This requires further research.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Tian Z, Ren Y, Wang G. Short-term wind power prediction based on empirical mode decomposition and improved extreme learning machine. J Electr Eng Technol 2018;13:1841–51.
[2] Maldonado-Correa J, Solano J, Rojas-Moncayo M. Wind power forecasting: A systematic literature review. Wind Eng 2019. 0309524X19891672.
[3] Singh SN, Erlich I. Strategies for wind power trading in competitive electricity markets. IEEE Trans Energy Convers 2008;23:249–56.
[4] Hong D, Ji T, Li M, Wu Q. Ultra-short-term forecast of wind speed and wind power based on morphological high frequency filter and double similarity search algorithm. Int J Electr Power Energy Syst 2019;104:868–79.
[5] Gangui Y, Yu L, Gang M, Yang C, Junhui L, Jigang L, Lei M. The ultra-short-term prediction of wind power based on chaotic time series. Energy Procedia 2012;17:1490–6.
[6] Zhang ZZ, Zou JX, Zheng G. Ultra-short term wind power prediction model based on modified grey model method for power control in wind farm. Wind Eng 2011;35:55–67.
[7] Shi K, Qiao Y, Zhao W, Wang Q, Liu M, Lu Z. An improved random forest model of short-term wind-power forecasting to enhance accuracy, efficiency, and robustness. Wind Energy 2018;21:1383–94.
[8] Lee J, Wang W, Harrou F, Sun Y. Wind power prediction using ensemble learning-based models. IEEE Access 2020;8:61517–27.
[9] Nielson J, Bhaganagar K, Meka R, Alaeddini A. Using atmospheric inputs for artificial neural networks to improve wind turbine power prediction. Energy 2020;190:116273.
[10] Li F, Ren G, Lee J. Multistep wind speed prediction based on turbulence intensity and hybrid deep neural networks. Energy Convers Manage 2019;186:306–22.
[11] Jacobson MZ, Archer CL. Saturation wind power potential and its implications for wind energy. Proc Natl Acad Sci 2012;109:15679–84.
[12] Kaiser K, Langreder W, Hohlen H, Højstrup J. Wind energy. Springer; 2007.
[13] Lubitz WD. Impact of ambient turbulence on performance of a small wind turbine. Renew Energy 2014;61:69–73.
[14] Türk M, Emeis S. The dependence of offshore turbulence intensity on wind speed. J Wind Eng Ind Aerodyn 2010;98:466–71.
[15] Trenkler G. Methods of multivariate analysis. Wiley series in probability and mathematical statistics: Probability and mathematical statistics section. Comput Statist Data Anal 1996;22:334–5.
[16] Samadianfard S, Hashemi S, Kargar K, Izadyar M, Mostafaeipour A, Mosavi A, Nabipour N, Shamshirb S. Wind speed prediction using a hybrid model of the multi-layer perceptron and whale optimization algorithm. Energy Rep 2020;6:1147–59.
[17] Zhao Y, Zhang Y. Comparison of decision tree methods for finding active objects. Adv Space Res 2008;41:1955–9.
[18] Ho TK. Random decision forests. In: IEEE proceedings of 3rd international conference on document analysis and recognition. 1995.
[19] Babar B, Luppino LT, Boström T, Anfinsen SN. Random forest regression for improved mapping of solar irradiance at high latitudes. Sol Energy 2020;198:81–92.
[20] Kleinbaum DG, Kupper LL, Nizam A, Rosenberg ES. Applied regression analysis and other multivariable methods. Nelson Education; 2013.