Available online at www.sciencedirect.com
ScienceDirect
Energy Reports 8 (2022) 916–922
www.elsevier.com/locate/egyr

2022 The 4th International Conference on Clean Energy and Electrical Systems (CEES 2022), 2–4 April, 2022, Tokyo, Japan

Noise-intensification data augmented machine learning for day-ahead wind power forecast

Hao Chen[a,c,*], Yngve Birkelund[b,c], Bjørn-Morten Batalden[a], Abbas Barabadi[a]

[a] *Department of Technology and Safety, UiT The Arctic University of Norway, Tromsø 9019, Norway*
[b] *Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø 9019, Norway*
[c] *Arctic Centre for Sustainable Energy, UiT The Arctic University of Norway, Tromsø 9019, Norway*

Received 14 May 2022; accepted 26 May 2022
Available online 10 June 2022

Abstract

The day-ahead wind power forecast is essential for the designation of dispatch schedules for the grid and rational arrangement for production planning by power generation companies. This paper specifically investigates the effect of adding noise to the original wind data for forecasting models. Linear regression, artificial neural networks, and adaptive boosting predictive models based on data-intensification white noise and uniform noise are evaluated in detail and their superiority over the original data-based models is compared. The results demonstrate that solely injecting noise into the dataset can statistically boost the performance of all forecasting models with learning algorithms. The findings of this study suggest a fresh perspective for developing wind power prediction models and carry certain wind energy engineering merits.

Peer-review under responsibility of the scientific committee of the 4th International Conference on Clean Energy and Electrical Systems, CEES, 2022.

Keywords: Machine learning; Noise; Power forecast; Statistical inference; Wind energy; Data science

1. Introduction

Wind energy is a renewable energy resource with optimal exploitation conditions and commercialization prospects [1]. The network operation of wind power generation is an essential path to achieve the large-scale development and utilization of wind energy [2]. Wind power is characterized by volatility, intermittency, and low energy density. These characters affect grid balance, which may profoundly endanger grid security [3].

An effective day-ahead wind power prediction is the foundation for operations, grid-connected wind parks and the dispatch of power systems including wind power [4].

Wind power prediction is categorized into physical and statistical models. The former is based on atmospheric dynamics and boundary layer meteorological theory, converting the weather model output into wind information

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

https://doi.org/10.1016/j.egyr.2022.05.265

Peer-review under responsibility of the scientific committee of the 4th International Conference on Clean Energy and Electrical Systems, CEES, 2022.

at the hub height of the wind turbine and considering the wake effects between turbines, and finally deriving the wind generator power from the power curve of the turbine [5]. The data-driven latter establishes an explicit or non-explicit relationship between historical information and power generation using the historical data, in effect viewing the forecasting as a time-series regression problem and fitting it by statistical or machine learning methods [6]. The statistical models are generally AutoRegressive (AR), Moving Average (MA), and their combinations and its variances, etc. Machine learning approaches, adaptively learning complex relationships between inputs and outputs, are kernel-based, networks-based, and ensemble models, etc.

The wind data resources for data-driven approaches are typically from wind masts and turbine hub measurements or highly accurate wind forecasts, all of which are documented deterministic data. However, there is some wind-data uncertainty that is caused by inherited indetermination, categorized into randomness and fuzziness, of wind power [7]. The mechanisms include the stochastic nature of wind volatility and intermittency and local effects such as wake and topography; the wind conditions undergo discontinuous and rapid changes; turbulence with diverse amplitude and frequency changes with time; and fuzziness of the wind speed-power conversion curve, especially at the power curve turning point, etc.

The uncertainty mentioned above statistically manifests itself in wind power forecasting as follows: first, a deterministic historical wind dataset does not necessarily represent the future conditions adequately; second, historical data are often limited, allowing the algorithms to memorize all samples in the training set when using machine learning for forecasting, thus resulting in overfitting.

Noise, generally representing uncertainty in data, is usually regarded as a unfavorable factor in the investigation of wind time series, and noise in the original series is filtered in many wind forecasting studies through signal processing techniques [8,9]. However, in computer science, researchers realized adding noise in a neural network, as a form of data augmentation. may lead to generalization improvements, which is like a regularization operation [10,11].

Inspired by data augmentation techniques, we inject noise to input wind data, including previous power, horizontal and vertical wind speed, of the training set to forecast day-ahead wind power with the linear regression with attributes selection and the random forests in the present study.

The remainder is organized as follows. Section 2 introduces the addition of noise and used statistics and algorithms. Section 3 presents the experiment procedure and data. The results of the model's performance and its analysis are shown in Section 4. Furthermore, a brief conclusion is demonstrated in Section 5.

2. Noise injection, statistics, and algorithms

2.1. Noise addition

Two types of stochastic number generators are used in this research to extend the size of the training set; one is the most commonly used noise, i.e., white noise, and the other can be considered extensive noise, but it contains a certain amount of information. The former is the noise, for which its power spectral density remains constant in its full-frequency domain. In particular, when a white noise's amplitude distribution follows a Gaussian distribution and its power spectral density is uniformly distributed, it is referred to as Gaussian white noise [12]. The latter has a uniform distribution in magnitude and is usually generated with the Mersenne twister algorithm in computer science [13]. Both types of noise allow users to define the range, $[a, b]$ where they are generated, the former with 99.7% of the values distribute within a range of three standard deviations cantered on the zero value, and the latter data with a uniform distribution are completely within the defined range.

An important technique to improve machine learning model performance in supervised learning is to ensure that the training set is as big and representative as possible. We, therefore, scale one-year training data, ninety percent of the annual data, to five years scale, i.e., from 7884 to 39,420 samples, by injecting noise into the historical wind data. The range for two types of noise is gradually incrementally defined as:

$$[a, b] = [-0.05n, 0.05n]\,;\, n = 2, 3, 4, 5 \tag{1}$$

2.2. Statistics

A whole annual wind data with the hourly temporal resolution, including measured wind speed & direction and power, of a wind power park in the Arctic, are used to establish forecasting models. The meteorological data are given by the operating company, and the Norwegian Water Resources and Energy Directorate (NVE) offers power data.

For a data population $\boldsymbol{X} = \{x_1, x_2, \ldots, x_n\}$, its statistical characteristics are expressed as follows:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2}$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{3}$$

$$\hat{\gamma}_3 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^3 / \sigma^3 \tag{4}$$

$$\hat{\gamma}_4 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^4 / \sigma^4 - 3 \tag{5}$$

where $\bar{x}$ is the average value and σ is the standard deviation of the population. $\hat{\gamma}_3$ denotes skewness that measures distribution shape *leans* to one side of the mean value. $\hat{\gamma}_4$ represents kurtosis, which shows how much does a distribution *peakedness* differs from the normal distribution.

For a time series $\{X_t,\ t \in T\}$, take $t,\ s \in T$, and define $\rho(t, s)$ as the AutoCorrelation coeFficient (ACF) [14], where μ_t is the mean and DX_t is the variance.

$$\rho(t, s) = \frac{E\,(X_t - \mu_t)\,(X_s - \mu_s)}{\sqrt{DX_t \cdot DX_s}} \tag{6}$$

The Bonferroni method for multiple comparisons is used to statistically check whether significant differences exist in predations of the same algorithmic models with different training sets. For the two tests, their hypotheses are similar. H_0: The mean differences are zero; H_a: At least one difference does not equal zero. The method for the difference confidence interval is expressed as:

$$\left(\bar{Y}_1 - \bar{Y}_2\right) \pm \frac{t_{n-k,1-\alpha/l}}{\sqrt{2}} \cdot \sqrt{MSE} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \tag{7}$$

where t is t distributions, k is the number of populations and n is the total size of all populations, l represents the number of pairs needed to be compared, and *MSE* is the Mean Square Error within groups.

2.3. Algorithms

This study conducts wind power multivariate regression prediction by employing three representative machine learning algorithms. The first one is Linear Regression with attribute selection (LR); the second one is Artificial Neural networks (ANN) and the last one is Adaptive Boosting (AdaBoost) in which linear regression with attributes selection is applied as its base learner. As this paper is not concerned with the details of these machine algorithms and Ref. [15] provides them in detail; only a brief description of these algorithms is given below.

LR: LR is a fundamental supervised machine learning algorithm due to its simplicity and well-known properties. It allows the use of the least-squares function or maximum likelihood estimation or even the learning approach, harnessed in this research, to determine the linear relationship that exists between the independent and dependent variables.

ANN: ANN represents a biomimetic intelligence algorithm that is inspired by the bio-neural. Multilayer Perceptrons (MLP) based on backpropagation optimization is the most popular ANN. It normally includes input, hidden, and output layers consisting of neurons. The number of neurons for input and outputs are based on the data structure and those for hidden layers are defined by users. These neurons are connected by weights and activated by various activation functions. The ANN structure of this paper consists of 19 neurons in the input layer, 30 in the hidden layer, 24 in the output layer, and the activation function is sigmoid.

AdaBoost with LR: Adaboost is a representative boosting ensemble learning algorithm. It constantly creates base learners, LR in the present study, to highlight (using larger weights) mislearning samples from previous learners until the number of learners hits a setting or the loss function meets a threshold. With regression issues, the weighted averaging is employed to acquire the eventual forecast results.

3. Experiments

Statistically, day-ahead wind power prediction can be treated as a multivariate regression problem, in which wind power series are autoregressive, and horizontal (West-East) and vertical (South-North) wind speeds supplement the information to this autoregression to enhance the forecast model. Since wind power is featured with daily similarity. [16] The similarity is more apparent for seaside wind farms because of the diurnal alternation of sea and land winds. Physically, therefore, wind information prior to 24 and its adjacent hours is potentially valuable for modeling current wind power. In this study, we chose data from two hours ante- and post-days as additional information for the next 24 h of multi-step prediction. Besides, to account for the effects of a substantial expansion of the sample size, a time coefficient term T is also included in the regression. The forecast as step $i + n$ is described as:

$$\hat{P}_{i+n} = f\left(P_{i-22}, \ldots, P_{i-26}; \boldsymbol{V}_{i-22}, \ldots, \boldsymbol{V}_{i-26}; T\right) + \varepsilon_n \tag{8}$$

where $\hat{P}_{i+n}$ is n time-step ahead forecasting wind power, $n \in \{0, 1, 2, \ldots, 23, 24\}$, $\boldsymbol{V}$ represents the wind velocity vector, u is defined as horizontal wind speed $u = |\boldsymbol{V}| \times \sin\theta$ and v is as vertical wind speed $v = |\boldsymbol{V}| \times \cos\theta$, θ is wind direction angle of $\boldsymbol{V}$, ε_n is the error of the predictive equation.

Two metrics were used in evaluating models for wind power forecasts. Namely, Root Mean Square Error (RMSE) and Mean Directional Accuracy (MDA). The first is based on the loss functions of our regression algorithms and the second is a predictive error direction indicator for indexing forecasting upward or downward directions.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(P - \hat{P})^2} \tag{9}$$

$$\mathrm{MDA} = \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{sgn\left(P-\hat{P}\right)} \tag{10}$$

where P is *observed* wind power, $\hat{P}$ is its predictive value, and $\mathbf{1}_{\mathrm{sgn}(.)}$ represents the indicator function.

For the original dataset, 90% are selected as the training set and the remaining 876 samples, accounting 10%, are used as the testing set. The noise is incrementally injected into the training set to establish new training sets, which are learned by the mentioned machine learning algorithms, and then the models are tested on the same testing set.

4. Results

4.1. Statistics of datasets

Table 1 shows the number of samples and the statistics, described above, for the original, white noise augmented, and uniform noise augmented datasets. The variables in these three sets are orderly shortened as O, N, and U separately, plus u, v, and P.

It is seen that the size of the dataset increases remarkably after the addition of noise, but their related statistics are not significantly different from the original dataset, indicating the statistical stability of the augmented datasets. All of the above statistics are even closer for the N-enhanced and U-enhanced datasets, with differences of less than 0.001. Specifically, the skewness and kurtosis of u and v show the historical distributions of the two speeds, are generally symmetric and have more concentrated peaks and longer tails than the normal distribution.

4.2. Forecasts

For day-ahead wind power forecasting, we develop LR, ANN, and AdaBoost models and perform 24-step, hourly, forecasts on the original, the white noise added and the uniform noise added datasets, respectively. These nine models are shortened as OLR, OANN, OAD; NLR NANN, NAD; and ULR, UANN, UUN; which correspond to the mentioned datasets in order.

Table 1. Statistics of datasets.

Statistics	Sample	$\bar{x}$ (m/s)	σ (m/s)	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\rho(1, 24)$
Ou	8760	0.122	6.327	0.055	1.065	−0.012
Nu	40296	0.134	6.189	0.071	1.025	−0.013
Uu	40296	0.134	6.191	0.071	1.024	−0.013
Ov	8760	0.003	6.292	−0.021	0.982	−0.019
Nv	40296	0.038	6.186	−0.015	0.998	−0.014
Uv	40296	0.037	6.187	−0.016	0.999	−0.015
OP	8760	15.239	15.858	1.020	−0.168	0.214
NP	40296	14.972	15.851	1.042	−0.124	0.226
UP	40296	14.972	15.852	1.041	−0.124	0.226

The RMSE growth related to 24 to 30 step forecast of models with original data compared with noise-enhanced data models are shown in Table 2. It is evident that simply injecting noise into the training set, both white and uniform, improves the performance of the original models by more than five percent. Where, the noise boosts the LR and AdaBoost forecasting models more significantly, with the RMSE reduced by more than 10%.

Table 2. RMSE growth related to 24 to 30 steps forecasts.

Growth (%)	LR		ANN		AD	
Steps	O v N	O v U	O v N	O v U	O v N	O v U
24	11.473	11.469	5.366	6.136	11.115	11.112
25	11.471	11.467	5.414	6.203	11.108	11.104
26	11.463	11.459	5.439	6.216	11.088	11.084
27	11.439	11.435	5.440	6.215	11.041	11.038
28	11.399	11.395	5.480	6.255	10.980	10.977
29	11.343	11.339	5.460	6.255	10.919	10.916
30	11.295	11.292	5.461	6.247	10.877	10.873
Average	11.412	11.241	5.437	6.230	11.018	10.838

The RMSE and MDA of multiple-step wind power forecasts are shown in Figs. 1 and 2. As seen in Fig. 1, all models' RMSE slowly increases with the forecast steps, but there is a jumping growth after 21 steps. The OLR and OAD models have the largest RMSE, and OANN has a larger RMSE, indicating the underperformance of all three prediction algorithms based on the original data. The RMSE of the models with the additional noise is significantly reduced, which implies a better fitting performance, with almost the same superior presentation for all models except the ANN-based models, which show a slightly poorer result.

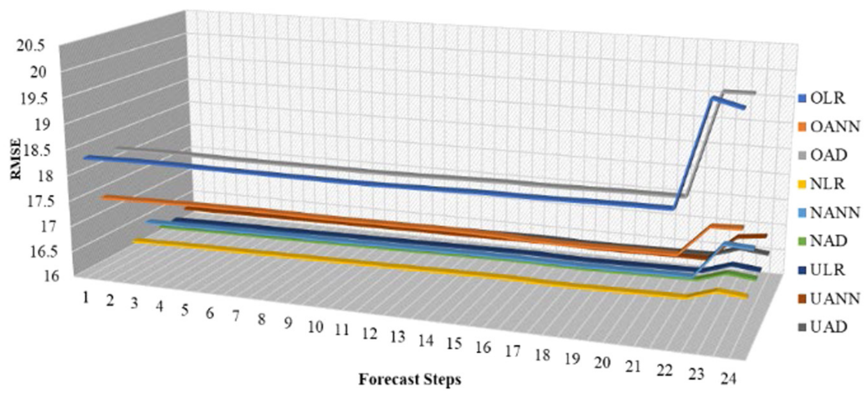

Fig. 1. The RMSE of wind power forecasts.

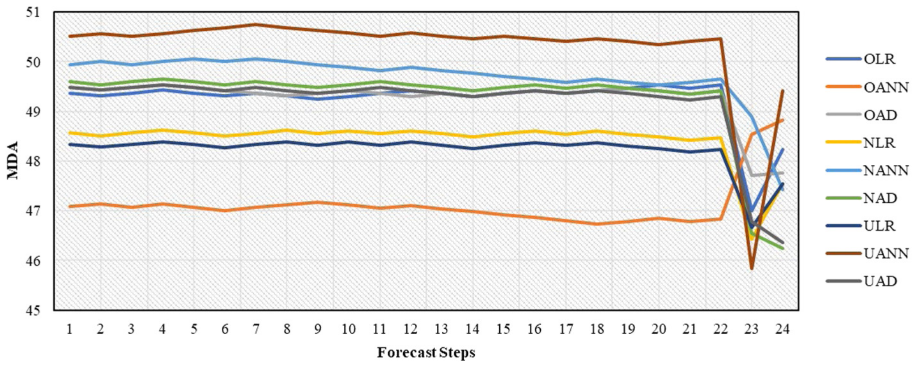

Fig. 2. The MDA of wind power forecasts.

Table 3. Model comparisons with Bonferroni method.

	RMSE			MDA		
Model comparison	p	Lower	Upper	p	Lower	Upper
OLR-NLR	0.000*	1.983	2.426	0.000*	0.486	1.145
OLR-ULR	0.000*	1.982	2.425	0.000*	0.689	1.347
NLR-ULR	1*	−0.222	0.220	0.409	−0.127	0.532
OANN-NANN	0.000*	0.795	1.041	0.000*	−3.055	−2.052
OANN-UANN	0.000*	0.944	1.190	0.000*	−3.663	−2.659
NANN-UANN	0.012	0.026	0.272	0.012	−1.109	−0.106
OAD-NAD	0.000*	1.904	2.356	1*	−0.520	0.533
OAD-UAD	0.000*	1.904	2.355	1*	−0.429	0.625
NAD-UAD	1*	−0.227	0.225	1*	−0.435	0.618

For MDA, a significant effect of noise on models' predicted direction could not be observed. Since MDA is only a counting indicator, through which the stability of the model can in one aspect be viewed. It is seen that all models show dramatic variations in MDA after 21 steps, possibly due to model collapses, which explains the rapid deterioration of RMSE appearing on the right side of Fig. 2. It implicitly means that after 21 steps these models can no longer reliably predict wind power.

Moreover, multiple pair comparisons of metrics with the Bonferroni method are conducted within the LR, ANN, and AdaBoost algorithms with different inputs. The model-comparison p-values and lower and upper bounds of for multiple-step RMSE and MDA are displayed in Table 3.

The p-values of RMSE comparisons reveal that the performance of new models with noise additives is statistically different from that of original models. Except for the ANN-based models, white noise and uniform noise have no significant effect on RMSE. RMSE difference bounds calculated by the Bonferroni method shows, in general, the noise-augmented LR and AdaBoost have a greater RMSE decrease than the ANN models. Specifically, UANN model has a statistically smaller RMSE than NANN, which indicates that uniform noise is more effective in our ANN. Regarding MDA, there is no obvious pattern for multiple comparisons. However, there are statistical influences of noise in LR and ANN forecasting algorithms. But whichever noise as a whole has no significant effect on MDA. In particular, the MDA of the AdaBoost-based model is immune to noise.

5. Conclusion

This paper systematically investigates two approaches of adding of noise to wind data and their manifestations in three machine learning-based multiple regression day-ahead wind power forecasting models. In accordance with the results, the following conclusions can be drawn.

First, multiple regression models based on linear regression, with attribute selection, neural networks, and adaptive boosting can all provide fairly stable and effective forecasts of day-ahead wind power within certain time steps. Second, merely adding noise to the training set boosts the performance of the predictive models, with these improvements exceeding 11.3, 5.5, and 10.9 percents on average for the linear regression, the neural networks, and the adaptive boosting, respectively. Third, the superior performance of the model based on data both white noise and uniform noise addition is statistically remarkable, which indicates that this method has certain generalizability and engineering applicability in other related fileds.

Future research may concentrate on developing approaches to incorporate noise injection into AI forecasting models and further improve predictive accuracy. The framework could also be extrapolated to cover other areas in energy research.

CRediT authorship contribution statement

Hao Chen: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Yngve Birkelund:** Comments on the manuscript. **Bjrn-Morten Batalden:** Comments on the manuscript. **Abbas Barabadi:** Comments on the manuscript.

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.

Data availability

The power data are publicly available while weather data are property of a local power company.

Acknowledgments

This study is supported by the Department of Technology and Safety and Arctic Centre for Sustainable Energy, UiT The Arctic University of Norway. H.C thanks Catalonia Institute for Energy Research (IREC) for its hosting for a 2022 spring academic visit, during which the paper was finalized.

References

[1] Zhixin W, Chuanwen J, Qian A, Chengmin W. The key technology of offshore wind farm and its new development in China. Renew Sustain Energy Rev 2009;13(1):216–22.
[2] Wu Y, Lau VK, Tsang DH, Qian LP, Meng L. Optimal energy scheduling for residential smart grid with centralized renewable energy source. IEEE Syst J 2013;8(2):562–76.
[3] Rahimi E, Rabiee A, Aghaei J, Muttaqi KM, Nezhad AE. On the management of wind power intermittency. Renew Sustain Energy Rev 2013;28:643–53.
[4] Shi X, Lei X, Huang Q, Huang S, Ren K, Hu Y. Hourly day-ahead wind power prediction using the hybrid model of variational model decomposition and long short-term memory. Energies 2018;11(11):3227.
[5] Kariniotakis G. Renewable energy forecasting: From models to applications. Woodhead Publishing; 2017.
[6] Maldonado-Correa J, Solano J, Rojas-Moncayo M. Wind power forecasting: A systematic literature review. Wind Eng 2019;0309524X19891672.
[7] Yusheng X, Xing L, Feng X, Chen Y, Zhaoyang D. A review on impacts of wind power uncertainties on power systems. Proc CSEE 2014;34(29):5029–40.
[8] Dong Q, Sun Y, Li P. A novel forecasting model based on a hybrid processing strategy and an optimized local linear fuzzy neural network to make wind power forecasting: A case study of wind farms in China. Renew Energy 2017;102:241–57.
[9] Du P, Wang J, Yang W, Niu T. A novel hybrid model for short-term wind power forecasting. Appl Soft Comput 2019;80:93–106.
[10] Badola A, Nair VP, Lal RP. An analysis of regularization methods in deep neural networks. In: 2020 IEEE 17th India council international conference. IEEE; 2020, p. 1–6.
[11] Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning (No. 2). Cambridge: MIT Press; 2016.
[12] Wang Y. Principles of communications. 2011.
[13] Matsumoto M, Nishimura T. Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Trans Model Comput Simul 1998;8(1):3–30.
[14] Wei WW. Time series analysis. In: The Oxford handbook of quantitative methods in psychology, vol. 2, 2006.
[15] Mohammed M, Khan MB, Bashier EBM. Machine learning: Algorithms and applications. CRC Press; 2016.
[16] Dong L, Wang L, Khahro SF, Gao S, Liao X. Wind power day-ahead prediction with cluster analysis of NWP. Renew Sustain Energy Rev 2016;60:1206–12.