Available online at www.sciencedirect.com
ScienceDirect
Energy Reports 8 (2022) 618–626
www.elsevier.com/locate/egyr

TMREES22-Fr, EURACA, 09 to 11 May 2022, Metz-Grand Est, France

A comprehensive statistical analysis for residuals of wind speed and direction from numerical weather prediction for wind energy

Hao Chen*

Department of Technology and Safety, Tromsø 9019, Norway
United Nations Conference on Trade and Development Unctad, Palais des Nations 1211 Geneva 10, Switzerland

Received 13 June 2022; accepted 18 July 2022
Available online xxxx

Abstract

Wind data are vital for the research in renewable energy research. Their quality from numerical weather prediction significantly influences the wind energy models. This paper utilizes a comprehensive statistical analysis for analyzing predictive errors, named residuals of wind speed and direction modeled by numerical weather prediction models. The analysis, taken an Arctic wind site as an example, effectively integrates statistical inference, probabilistic modeling, and hypothesis tests. It is proven that the residuals still contain important meteorological information. The introduced statistical analysis may be used to replenish residuals and explore complex intrinsic properties of numerical weather wind models and contributions to wind energy modeling.

Peer-review under responsibility of the scientific committee of the TMREES22-Fr, EURACA, 2022.

Keywords: Numerical weather prediction; Residual analysis; Statistical modeling; Hypothesis test; Wind energy

1. Introduction

Wind is both the principal object and an important approach to renewable energy science research [1–3]. As one of the most fundamental natural phenomena, wind modeling is the most important driver that approaches the atmosphere [4], research, and serves as a crucial sustainable energy resource assessment [5,6]. The effective wind model for a site should understand the historical wind characteristics and be able to predict the wind temporally based on these characteristics. One of the most effective ways to conduct a comprehensive environmental and energy assessment of regional wind is to construct a target candidate site wind model based on long-term wind data measured by weather towers. However, in practice, it is not always possible to make weather measurements at candidate sites, and the installations and operations of these towers are quite expensive [7]. Meanwhile, Numerical Weather Prediction (NWP), which can derive wind climatology with different resolutions at regional scales, is regarded as an important alternative source of wind data [8].

* Correspondence to: Department of Technology and Safety, Tromsø 9019, Norway.
E-mail address: hao.chen@uit.no.

https://doi.org/10.1016/j.egyr.2022.07.080

Peer-review under responsibility of the scientific committee of the TMREES22-Fr, EURACA, 2022.

Most of the research on using statistical methods to evaluate wind data is based on probabilistic function modeling [9,10]. However, the majority of these studies focused on evaluations for wind speed distribution [11,12] and ignored wind directions.

Some studies are concerned with utilizing statistical analysis to investigate the forecast errors (named *residuals* in regression analysis). The majority of these studies look at probabilistic modeling for predictive errors but not to predict the residuals. Some research only looked at the normal distribution, which has proven not as accurate as other more appropriate distributions. M. Lange et al. [13] analyzed the uncertainty in wind modeling using statistical distributions and found that the error is normally distributed. J. Wu et al. [14] used a mixed distribution to model the error of the persistence model of wind energy. H. Wang et al. [15] investigated the normal distribution and kernel density estimation to model wind speed error between NWP and measured wind speed data and found that the mapping relationship is vague. The critical reason is the existing correction algorithms in NWP algorithms. Some tried to use different distributions to analyze the problem. P. Guo et al. [16] analyzed the fluctuation in wind direction by using the Weibull distribution to fit the marginal probability density of fluctuation amplitude and fluctuation duration and combined them with the mixed Copula and proved the accuracy by testing the model in a wind farm in China. N. Chen et al. [17] used the Gaussian process to correct wind speed data from NWP and demonstrated its edge by employing the correct wind data in two wind site models.

This study elaborates on the statistics, time series, and quarterly probability distribution properties of the NWP wind model residuals. It provides a new perspective for further in-depth investigations and optimizations of the complex numerical weather systems for wind power engineering. Besides, the paper conducts several hypothesis tests applied in statistical modeling and machine learning. These statistical tests are usually missed in common energy engineering studies. Moreover, we especially present a detailed literature review of this minority research field of wind residual analysis in Section 1, which can help researchers follow the field more conveniently.

The remainder of this paper is organized as follows. Section 2 shows the statistical methods. In Section 3, the corresponding case study setup is introduced. Section 4 elaborates on the results of the proposed framework. The main conclusions are briefly presented in Section 5.

2. Methodology

In this section, the theoretical methods involved in statistical analysis are demonstrated.

Residual analysis is a crucial part of statistical regression diagnostics techniques. The residual $\hat{E}$ is defined in (1):

$$\hat{E} = Y - \hat{Y} \tag{1}$$

where Y is the observation and $\hat{Y}$ is the regression value from predictive models. There are few assumptions for the residual, with random and unpredictable characteristics, in linear regression. The first is residuals are independent of the data sample itself. The second is residuals are independent of each other and have the same probability distribution. The third is residuals should follow the standard normal distribution [18]. If the residuals do not meet these three assumptions, the regression model does not correctly exploit the data's information, and there is still space for improving the model.

The probability density function (PDF) of a random variable is a mathematical model that describes the probability of this variable happens at a particular point in each observation interval. The cumulative distribution function (CDF) indicates the possibility that a variable is less than or equal to a specific value [19]. In this study, we will use four PDF ideal distributions, namely, normal distribution, skew normal distribution, t distribution, and stable distribution.

For the normal distribution, its PDF is expressed by (2):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{2}$$

where μ is the mean and σ is the standard division.

The skew normal distribution is an extended normal distribution, it adds nonzero skewness into the distribution [20]. Its PDF is shown in (3):

$$f(x) = \frac{2}{\omega\sqrt{2\pi}} e^{-\frac{(x-\xi)^2}{2\omega^2}} \int_{-\infty}^{\alpha\left(\frac{x-\xi}{\omega}\right)} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt, \sigma = \xi + \omega\delta\sqrt{\frac{2}{\pi}}, s^2 = \omega^2\left(1 - \frac{2\delta^2}{\pi}\right), \text{ where } \delta = \frac{\alpha}{\sqrt{1+\alpha^2}} \tag{3}$$

where ω is a scale parameter, ξ is a location parameter and α is a shape parameter.

The PDF of t distribution is determined via the following function [21]:

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\nu\pi}} \frac{1}{\left(1 + \frac{x^2}{\nu}\right)^{\frac{\nu+1}{2}}} \tag{4}$$

where $\nu > 0$ is the degrees of freedom and $\Gamma(.)$ is the Gamma function.

Stable distributions are a class of probability distributions suitable for modeling heavy tails and skewness [22]. The function determines the PDF of stable distribution in (5):

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi(t) e^{-ixt} dt \tag{5}$$

where $\varphi(t)$ is expressed by (6):

$$\varphi(t) = \begin{cases} exp\left(-\gamma^\alpha |t|^\alpha \left[1 + i\beta\ sign(t) \tan\frac{\pi\alpha}{2}\left((\gamma|t|)^{1-\alpha} - 1\right)\right] + i\delta_0 t\right) & for\ \alpha \neq 1 \\ exp\left(-\gamma|t|\left[1 + i\beta\ sign(t)\frac{2}{\pi}\ln(\gamma|t|)\right] + i\delta_0 t\right) & for\ \alpha = 1 \end{cases} \tag{6}$$

where α is a first shape parameter and $0 < \alpha \leq 2$, β is a second shape parameter and $-1 \leq \beta \leq 1$, γ is a scale parameter and $0 < \gamma < \infty$, δ is a location parameter and $-\infty < \delta < \infty$. α describes the tails of the distribution. β represents the skewness of the distribution. Specifically, when α equals 2, the stable distribution is the normal distribution.

A nonparametric estimated PDF fitting approach based on histograms, called kernel distribution, is introduced to examine the performance of ideal distributions for formulating. It is a smoothing technique that makes the discontinuous histograms into a kind of continuous PDF curve. It is defined by a smoothing function $K(\cdot)$ and a bandwidth d in (7):

$$\hat{f}_d(x) = \frac{1}{nd} \sum_{i=1}^{n} K\left(\frac{x - x_i}{d}\right) \tag{7}$$

To determine the distributions' parameters for a given dataset, the Maximum Likelihood Estimation (MLE) method is used for the parametric estimations for different PDFs. Under the MLE criterion, a sample is considered from an aggregate that maximizes that particular sample's probability.

Then, we introduce some nonparametric hypothesis testing for checking the statistical significance.

Normality Test: The Anderson–Darling test is a statistical hypothesis test of whether the sample data are from a normal distribution [23]. It is one of the most powerful statistical tools for detecting normality.

H_0: The data follow a normal distribution.

H_a: The data do not follow a normal distribution.

The test statistic is given as in (8) [23]:

$$A^2 = -n - S, S = \sum_{i=1}^{n} \frac{2i-1}{n} \left[ln\left(F\left(X_i\right)\right) + ln\left(1 - F\left(X_{n+1-i}\right)\right)\right] \tag{8}$$

where $\{X_1 < \cdots < X_n\}$ is in order, $F(.)$ is CDF of the normal distribution. The test statistic can be compared to critical values from the theoretical distribution.

Autocorrelation Test: The Ljung–Box test is a statistical test to check for the presence of autocorrelation in a series [24]. It is a powerful portmanteau test since it tests overall randomness according to fixed multiple lags rather than tests randomness based on each lag.

H_0: The data are distributed independently.

H_a: The data are not distributed independently.

The test statistic is defined as in (9) [24]:

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k} \tag{9}$$

follows $\chi^2_{(h)}$ under H_0, where $\hat{\rho}_k$ is autocorrelation at lag k, h is the testing lags number.

Stationarity Test: The Augmented Dickey–Fuller test (ADF) tests the existence of unit root for checking stationarity in a series [25]. The unit root that equals one refers to a feature that makes a time series non-stationary.

H_0: The unit root equals one; the data are non-stationary.

H_a: The unit root is smaller than one; the data are stationary.

Goodness-of-fit Test: The one-sample Kolmogorov–Smirnov test (K–S test) is a statistical test based on CDF to test whether a distribution is from a kind of ideal distribution [26]. Moreover, for two datasets, if the ideal distribution is replaced with the other dataset real distribution, a one-sample K–S test can be extended to a two-sample K–S test, which tests whether the two datasets come from the same distribution.

H_0: The data have a given distribution.

H_a: At least one value does not come from the given distribution.

The test statistic is defined as in (10) [26]:

$$D = \sup_x |F_0(x) - F(x)| \tag{10}$$

where $F_0(x)$ is CDF of the given ideal distribution, and $F(x)$ is CDF of the testing data. The test statistic can be compared to critical values from the theoretical distribution.

Rank Test: The Wilcoxon signed-rank test is a paired difference test to assess whether the two populations' medians differ [27].

H_0: The two populations have the same median.

H_a: The two populations have different medians.

There is an implementation of ADF and Wilcoxon signed-rank tests in *R* language.

The Friedman test, similar to analysis of variance (ANOVA), is used to check for differences in performance across multiple trials [27]. Especially, it tests for column effects after adjusting for possible row effects with the *F* statistic.

H_0: The column data do not have a significant difference.

H_a: The column data have a significant difference.

Performance Evaluation Metrics for PDF Modeling: Three metrics are used to evaluate the wind velocity residuals modeled with ideal PDFs. Namely, Root Mean Square Error (RMSE), coefficient of determination (R^2), and *p*-value of one-sample K–S test. We use them to calculate the probability density difference between smoothing PDFs (kernel PDFs) and corresponding ideal parametric distributions. In the one-sample K–S test, the *p*-value is the probability that the data fits the given distribution on the extreme conditions.

3. Experimental setup

A Norwegian Arctic wind park gives the measured wind speed and direction data. The wind data was modeled, 2.5 km resolution, by the Scandinavian weather institutions (an NWP model named Meteorological cooperation on operational Ensemble Prediction System MEPS). The timestamp of the two datasets is from 0:00 1st January 2017 to 23:00 31st December 2017 with a one-hour resolution. The data are divided into 4-quarters datasets with numbers of 2160, 2184, 2208, and 2208.

A wind speed is a scalar value for the wind velocity vector. East–West wind speed and North–South wind speed are the East–West and North–South scalar values of the wind velocity vector (including information of vector's scale and direction). The East–West wind speed (u), North–South wind speed (v), and wind velocity vector ($\boldsymbol{V}$) are expressed in (11):

$$u = p \times \sin\theta;\ v = p \times \cos\theta;\ \boldsymbol{V} = \{p, u, v\} \tag{11}$$

The NWP model residuals are in definition as the difference between measured and NWP predicted wind velocities. Their abbreviations are *P, u, v* (measured overall, East-West and North-South wind speed); P_N, u_N, v_N

(overall, East-West and North-South wind speed calculated by NWP model); *RP, Ru, Rv* (residual overall, East-West, and North-South wind speed calculated by measured data minus correspond NWP data).

The whole experiment can be divided into two sections. First, the descriptive statistics are calculated for the overall, East-West, and North-South wind speed NWP data along with the corresponding actual measured data. Performing these homogeneous distribution tests for these two datasets and calculating the residuals of the wind velocity from the NWP model. The second part is statistical analysis for residuals of wind velocity. We apply rigorous statistical inference and time series testing techniques to analyze the residual series. The residuals' historical distributions are also modeled with PDFs. A comparative assessment is made to identify an ideal distribution model that is most appropriate for describing the residuals' probabilistic characteristics. The experimental process is shown is Fig. 1.

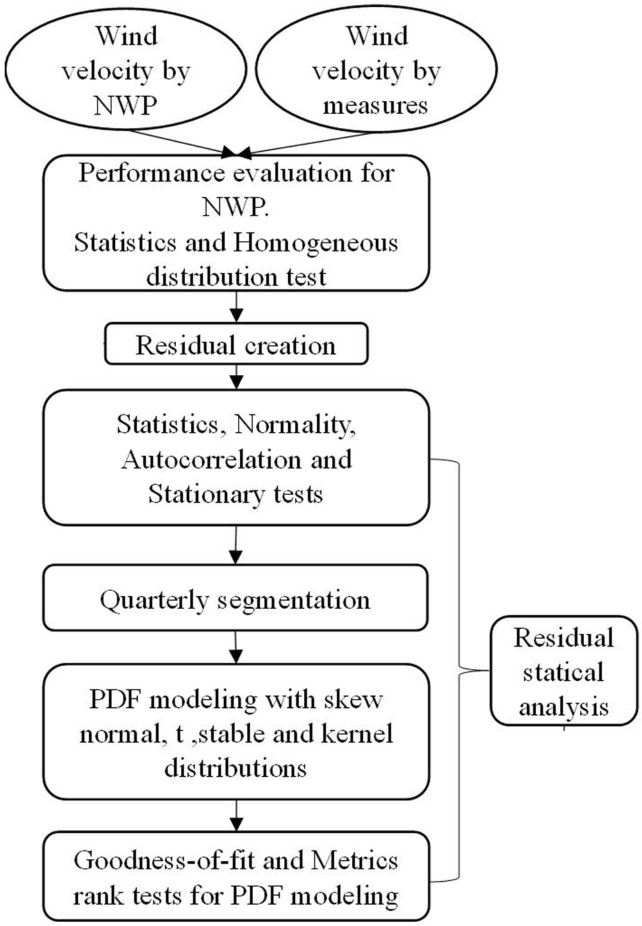

Fig. 1. The experimental procedure for the statistical analysis.

4. Results and discussions

4.1. NWP for wind velocity

The descriptive statistics of the wind velocity from real measurements and the NWP model are shown in Table 1. It can be seen that the actual measured annual mean wind speed is 7.69 m/s, which is 0.74 m/s larger than the annual average of predicted wind speed from NWP. The wind velocity from NWP is less volatile and has a smaller value range than the measured one. The P and P_N have positive skewness that indicates their right tails are longer than the left ones, and the mass of their distributions are concentrated on the left. The skewness of u, u_N, v, and v_N

Table 1. Statistics of wind velocity.

Statistics	P	P_N	u	u_N	v	v_N
Mean (m/s)	7.69	6.95	0.12	0.33	0.00	1.29
Standard Deviation (m/s)	4.53	3.88	6.33	5.25	6.29	5.83
Max (m/s)	33.88	33.69	32.68	25.57	33.11	19.39
Min (m/s)	0.10	0.10	−33.16	−27.89	−31.51	−18.63
Skewness	1.34	1.15	0.05	0.07	−0.02	−0.24
Kurtosis	2.66	2.96	1.07	1.81	0.98	−0.54

approximately equal to zero, indicating that their probability distribution has some symmetry. P and PN's kurtosis is significantly greater than zero, suggesting that they are leptokurtic and have steeper or thicker tails than the normal distribution. Meanwhile, u, u_N, v, and v_N have slightly different kurtosis from zero, illustrating that the normal distribution cannot characterize them well.

Table 2. Performance of persistence and NWP model.

Metrics	P	u	v
Persistence RMSE	1.5092	8.72	8.7578
NWP RMSE	2.949	8.2028	8.6885
Persistence 1-R^2	0.1079	0.9975	0.9990
NWP 1-R^2	0.4053	1	0.9999

The two-sample K–S and Wilcoxon signed-rank tests are conducted between the two wind velocity datasets. Their p-values are all but entirely zero, which shows that the two datasets are significantly different in probability distributions. These can also be seen in the differences in their descriptive statistics. The forecasting performance of NWP and persistence model, in which the value of t+1 equals the value of t, for overall, East–West, and North–South wind speeds is shown in Table 2. It is seen that for the prediction of P, the RMSE and 1-R^2 of the NWP model both exceed those of the persistence model, explaining a better performance by the persistence model. The reason is that the NWP is a mesoscale weather model, while the persistence model inputs are real measured wind speeds at a previous time, but real wind speed data are not widely available in practice. However, by factoring the wind direction into prediction models, the NWP and persistence models show similar performance, indicating that the NWP model has a reasonable recognition of wind characteristics.

4.2. Statistical analysis of wind velocity residuals

The descriptive statistics of residual wind velocity are shown in Table 3. It is observed that the mean values of three residuals are close to zero, suggesting that errors from the NWP model are not one-sided, i.e., the predicted values oscillate around the actual values. The East–West and North–South wind speed residuals have greater volatility and range, indicating interactions between wind speed and direction errors. Skewness illustrates that they have a certain symmetry. Kurtosis shows they are leptokurtic and lightly steeper or thicker tailed than the normal distribution.

The time series tests are conducted to statistically check normality, autocorrelation, and stability of residual series with Anderson–Darling, Ljung–Box, and ADF tests, respectively. Their p-values are displayed in Table 4. Based on these p-values, we reject the null hypothesis in all cases with a confidence level of 0.01. This means that the residual series do not significantly follow a normal distribution, have some autocorrelation, and are stationary time series.

Concerning the residual probabilistic distribution analysis, their histograms are plotted and histograms-based kernel distribution fittings are applied with a smoothing normal distribution function and a bandwidth of 0.25 m/s. Besides, three different ideal distributions, established by MLE methods, are employed to model the residuals and their modeling performance is compared to the corresponding kernel distributions. Moreover, to further detect the original residual data probability distributions, we take one-sample K–S tests based on parametric ideal distributions and calculate their p-values. The modeling performance is displayed in Table 5.

Table 3. Statistics of residual wind velocity.

Statistics	Mean (m/s)	Standard Deviation (m/s)	Max (m/s)	Min (m/s)	Skewness	Kurtosis
RP	0.76	2.87	14.78	−12.09	0.30	1.38
Ru	−0.21	8.25	59.05	−45.12	0.31	3.26
Rv	−1.29	8.64	36.50	−49.35	−0.43	1.43

Table 4. *p*-values of time series tests.

Test	Anderson–Darling	Ljung–Box	ADF
RP	<0.0001	<0.0001	0.001
Ru	<0.0001	<0.0001	0.001
Rv	<0.0001	<0.0001	0.001

Table 5. Performance of PDF modeling for residuals.

Metrics	RMSE			1-R^2			*p*-value for K–S test		
Distribution	*RP*	*RU*	*RV*	*RP*	*RU*	*RV*	*RP*	*RU*	*RV*
Skew normal	0.0081	0.0046	0.0046	0.0237	0.0646	0.068	<0.0001	<0.0001	<0.0001
t	0.0039	0.0015	0.0039	0.0061	0.0075	0.0521	0.2633	0.2645	<0.0001
Stable	0.0042	0.0019	0.0038	0.0069	0.0118	0.0492	0.1398	0.0391	<0.0001

In PDF curves perspective, all the three PDFs can describe the main characteristics of the overall, East–West and North–South wind speed residuals historical probability bar charts, which have the shape of center-concentrated and symmetrical along the center, and it decays rapidly from the center and has thick tails. Besides, the figure of North–South residual is right-skewed. From Table 5, *t* and stable distributions have the lowest and similar RMSE and 1-R^2. Furthermore, both of them are not rejected by the null hypothesis of K–S tests with a confidence level of 0.01. These show that both distributions can describe the probability distribution of wind velocity prediction residuals with strict statistical significance. North–South residuals' right skewness also reduces the accuracy of ideal distributions in modeling them. Meanwhile, the skew normal distribution can still capture features of smoothing PDFs for original residual data since there are no remarkable performance differences between it and the other two distributions.

To further understand the temporal characteristics of the NWP wind velocity models and their PDF formulating, the annual wind velocity residual series is divided into four quarterly series. Namely, Q1, Q2, Q3, and Q4. And they are marked with 1, 2, 3, and 4 behind the residual abbreviations. Like the annual data modeling process, the PDF modeling of four quarters of residual data is also made. The quarterly PDF formulation performs similarly to the annual modeling. In particular, skew normal distribution is better than its counterpart in the annual analysis. Besides, more cases pass homogeneous distribution tests with a confidence level of 0.01 and some of them have large *p*-values.

The Friedman test is employed to carry out a rank test to compare the PDF formulating performance of residual data from different quarters and types of wind speed more objectively. We are interested in two effects. Is there any significant difference between the three ideal distributions models for a single type of wind speed over four quarters? The other is whether there is a significant difference between the different types of wind speed residuals modelings across quarters for the average performance of three PDFs? The Friedman test results for these two questions are displayed in Tables 6 and 7. All *p*-values surpass the confidence level of 0.01, so the Friedman test's null hypothesis should not be rejected. It is concluded that these three PDFs show differences in modeling the residuals across the year. The overall *p*-values of *RP*, *Ru,* and *Rv* ascent are in order, which indicates three PDFs' performance also vary for different wind speed residual formulations.

Similar to the results in Table 6, we reject all the Friedman tests a null hypothesis and confirm the significant difference in three PDFs' average performance between *RP*, *Ru,* and *Rv* annually. Notably, the average *p*-value of Table 6 is larger than the one of Table 7, which means the difference, in PDF formulating, between residuals of wind speed types is more significant than the one between different ideal fitting distributions.

Table 6. *p*-values of Friedman test for quarter with distributions effect.

Metrics	RMSE	1-R^2	*p*-value for K–S test
RP	0.0183	0.0183	0.0183
Ru	0.0381	0.0498	0.0498
Rv	0.3679	0.3679	0.2574

Table 7. *p*-values of Friedman test for quarter with wind speed effect.

Metrics	Q1	Q2	Q3	Q4
RMSE	0.0498	0.097	0.7165	0.7165
1-R^2	0.0498	0.0498	0.097	0.2636
K–S test *p*-value	0.097	0.0498	0.0859	0.5292

5. Conclusion

This paper focuses on a crucial issue in wind power generation prediction, namely the significant discrepancy between the numerical weather data, as inputs to the hybrid power forecast model for a wind site, and actual wind conditions at the site. These inputs error data severely affect the accuracy of the wind energy-related model. We construct a statistical analysis and learning prediction framework based on regression diagnosis for the NWP wind model itself. From the results in Section 4, the following conclusions can be drawn.

In general, the Scandinavian mesoscale NWP model achieves fairly accurate wind speed forecasts for the Arctic wind site. However, its performance degrades when joint forecasts of wind speed and direction are taken into account.

The analysis of the overall, East-West, and North-South NWP wind speed residual series based on statistics, quarterly probability modeling, and hypothesis test reveals that these series still contain valuable wind information that can be extracted by our proposed strict statistical modeling. Besides, these wind residuals are analogous to wind itself, which exhibits some quarterly volatility.

In further research, we hope to combine the proposed wind residuals analytical framework with the NWP wind model with higher resolution to obtain more accurate wind data for inputs of the superior performing wind energy models.

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

Data will be made available on request.

Acknowledgments

This study is financially supported by the Department of Technology and Safety and Arctic Centre for Sustainable Energy, Norway, UiT The Arctic University of Norway. The author thanks Dr. Yngve Birkelund for organizing the data. The contents expressed herein are those of the author and do not necessarily reflect the views of the United Nations.

References

[1] Bahaghighat M, Abedini F, Xin Q, Zanjireh MM, Mirjalili S. Using machine learning and computer vision to estimate the angular velocity of wind turbines in smart grids remotely. Energy Rep 2021;7:8561–76.
[2] Wang Y, Lu Z, Ma Y, Li X, Ren W, Zou X. Analysis on the characteristics of wind speed and wind energy resources from 1961 to 2020 and the impact of urban underlying surface change on them in shenyang. Energy Rep 2022;8:335–42.
[3] Chen H, Birkelund Y, Zhang Q. Data-augmented sequential deep learning for wind power forecasting. Energy Convers Manage 2021;248:114790.

[4] Gill TE, Shao Y. Introduction: Modelling of wind erosion and aeolian processes. Environ Modell Softw 2004;2(19):91–2.
[5] Rafiee A, Van der Male P, Dias E, Scholten H. Developing a wind turbine planning platform: Integration of sound propagation model–GIS-game engine triplet. Environ Model Softw 2017;95:326–43.
[6] Milashuk S, Crane WA. Wind speed prediction accuracy and expected errors of RANS equations in low relief inland terrain for wind resource assessment purposes. Environ Model Softw 2011;26(4):429–33.
[7] Jang Y, Byon E. Probabilistic characterization of wind diurnal variability for wind resource assessment. IEEE Trans Sustain Energy 2020.
[8] Carvalho D, Rocha A, Gómez-Gesteira M, Santos C. A sensitivity study of the WRF model in wind simulation for an area of high wind energy. Environ Model Softw 2012;33:23–34.
[9] Jung C, Schindler D. Wind speed distribution selection–A review of recent development and progress. Renew Sustain Energy Rev 2019;114:109290.
[10] Chen H, Anfinsen SN, Birkelund Y, Yuan F. Probability distributions for wind speed volatility characteristics: A case study of northern Norway. Energy Rep 2021;7:248–55.
[11] Aries N, Boudia SM, Ounis H. Deep assessment of wind speed distribution models: A case study of four sites in Algeria. Energy Convers Manage 2018;155:78–90.
[12] Alavi O, Mohammadi K, Mostafaeipour A. Evaluating the suitability of wind speed probability distribution models: A case of study of east and southeast parts of Iran. Energy Convers Manage 2016;119:101–8.
[13] Lange M. On the uncertainty of wind power predictions—Analysis of the forecast accuracy and statistical distribution of errors. J. Sol. Energy Eng. 2005;127(2):177–84.
[14] Wu J, Zhang B, Li H, Li Z, Chen Y, Miao X. Statistical distribution for wind power forecast error and its application to determine optimal size of energy storage system. Int J Electr Power Energy Syst 2014;55:100–7.
[15] Wang H, Han S, Liu Y, Yan J, Li L. Sequence transfer correction algorithm for numerical weather prediction wind speed and its application in a wind power forecasting system. Appl Energy 2019;237:1–10.
[16] Guo P, Chen S, Chu J, Infield D. Wind direction fluctuation analysis for wind turbines. Renew Energy 2020;162:1026–35.
[17] Chen N, Qian Z, Nabney IT, Meng X. Wind power forecasts using Gaussian processes and numerical weather prediction. IEEE Trans Power Syst 2013;29(2):656–65.
[18] Powers D, Xie Y. Statistical methods for categorical data analysis. Emerald Group Publishing; 2008.
[19] Pak A. Statistical inference for the parameter of lindley distribution based on fuzzy data. Braz J Probab Stat 2017;31(3):502–15.
[20] Azzalini A, Capitanio A. Statistical applications of the multivariate skew normal distribution. J R Stat Soc Ser B Stat Methodol 1999;61(3):579–602.
[21] Rigby RA, Stasinopoulos DM. Generalized additive models for location scale and shape. J R Stat Soc Ser C (Applied Statistics) 2005;54(3):507–54.
[22] Nolan JP. Stable distributions, math/stat department. American University; 2009.
[23] Anderson TW, Darling DA. Asymptotic theory of certain goodness of fit criteria based on stochastic processes. Ann Math Stat 1952;193–212.
[24] Ljung GM, Box GE. On a measure of lack of fit in time series models. Biometrika 1978;65(2):297–303.
[25] Greene WH. Econometric analysis. Pearson Education India; 2003.
[26] Justel A, Peña D, Zamar R. A multivariate Kolmogorov–Smirnov test of goodness of fit. Statist Probab Lett 1997;35(3):251–9.
[27] Gibbons JD, Chakraborti S. Nonparametric statistical inference: revised and expanded. CRC Press; 2014.