UiT The Arctic University of Norway

Faculty of Science and Technology
Department of Technology and Safety

Data-driven Arctic wind energy analysis by statistical and machine learning approaches

Hao Chen

A dissertation for the degree of Philosophiae Doctor – July 2022

Τα είς ἑαυτόν

Hårde tider har vi døyet, ble til sist forstøtt

Dedicated to existence

君を飾る花を咲かそう

想淵明《停雲》詩就，此時風味。

UiT The Arctic University of Norway

Data-driven Arctic wind energy analysis by statistical and machine learning approaches

The Ph.D. thesis is a collection of published or finished papers during the Ph.D. program.

Hao Chen

The thesis is the fulfillment of the partial requirements for the degree of Philosophiae Doctor (Ph.D.) in Natural Science at the UiT The Arctic University of Norway.

July 2022

Department of Technology and Safety, Faculty of Science and Technology;
Arctic Centre for Sustainable Energy;
UiT The Arctic University of Norway.

Acknowledgment

The investigations are the results of my doctoral studies from October 2019 to June 2022 in the Department of Technology and Safety. The study period is under the compliant supervision of Dr. Yngve Birkelund, Dr. Reidar Staupe-Delgado, Dr. Fuqing Yuan, and Dr. Stian Normann Anfinsen (first two years).

First, special thanks to Yngve Birkelund for his thoughtful guidance, persistent support, and invaluable comments. Needless to say, without his kindness in daily life and the freedom and resources he offered for my research, my Ph.D. work would not have been established into the current content. I am sincerely grateful to Reidar Staupe-Delgado for his encouragement and useful instructions for my work and career. I thank Fuqing Yuan and Stian Normann Anfinsen for their timely and insightful comments and suggestions on my work.

I am sincerely thankful to the Department of Technology and Safety for hiring me as a Ph.D. researcher and for enrolling me in a warm large family and administrative support and other Ph.D. students in the department for their accompany.

Thanks to the hosting supervisors for my academic visits or internship during my Ph.D.: Dr. Xiuhua Zhu at the Max Planck Institute for Meteorology, Hamburg, Germany; Dr. Cristina Corchero Garcia at the Catalonia Energy Research Institute, Barcelona Spain; Dr. Muhammad Aziz at the University of Tokyo, Tokyo, Japan; and Dr. Clovis Freire Junior at the United Nations Conference on Trade and Development (UNCTAD); Geneva, Switzerland.

I am grateful for the horror of being a member of the Arctic Centre for Sustainable Energy (ARC) UiT and the Norwegian research school on Changing Climates in the coupled Earth System (CHESS) UiB.

I am very grateful to the coauthors, editors, and reviewers of my publishing.

Last but not the least, I want to thank my family and friends for their unbounded love, support, and care during the period.

Abstract

Norway's Arctic region is rich in wind resources and developing wind energy in the region can promote a green transition and economic development. However, the region's unique topography with fjords and mountains and cold climate conditions make wind resource assessment, generation analysis, and power forecasting particularly challenging.

The accumulation of wind data and the emergence of data science give new promise to this issue. “Can advanced statistical and machine learning methods deliver effective and accurate analysis for wind energy in these Arctic landscapes that are characteristics with dramatically fluctuating wind?” The thesis systemically answers the question with the chronological order of the wind power generation process.

First, a statistical probabilistic modeling approach is utilized to assess wind energy resources in particular wind speed and its volatility, both from measured and numerically modeled wind data. The accurate assessment results contribute to evaluating wind resources of sites in the Arctic region.

Then, we propose a wind power curve model to monitor wind power generation for the Arctic wind park. The model involves quantifying wind turbulence, clustering meteorological data, and ensemble learning and reaching a satisfactory modeling result for the park power curve.

Finally, we demonstrate that traditional machine learning methods can be used to make short-term wind power forecasts for the Arctic wind parks, and these forecasts could be improved to some extent by applying appropriate meteorological wind data, as inputs, to the forecasting models. Moreover, we developed a novel approach for turbine forecasting with appropriate data processing techniques, and loading the data into large deep learning models allows for more accurate forecasting in different terrain conditions. Further, we utilized a variety of transfer learning techniques to make it possible to refine the raw data information and transfer large accurate but slow training forecasting models to smaller and faster ones for realizing rapid and efficient wind power forecasting.

In summary, through the above three parts of the investigation, the Ph.D. project achieved the target goal of developing data-driven Arctic wind energy analysis by statistical and learning approaches for wind parks and turbines in the Norwegian Arctic area.

List of appended scientific publications

No.	Authors	Publications
Paper I	Chen, Hao; Birkelund, Yngve; Anfinsen, Stian Normann; Staupe-Delgado, Reidar; Yuan, Fuqing	Assessing probabilistic modeling for wind speed from numerical weather prediction model and observation in the Arctic. nature Scientific Reports 2021; Volum 11. (7613) s.
	Chen, Hao; Anfinsen, Stian Normann; Birkelund, Yngve; Yuan, Fuqing	Probability distributions for wind speed volatility characteristics: A case study of Northern Norway. Energy Reports 2021; Volum 7. s. 248-255.
Paper II	Chen, Hao	Cluster-based ensemble learning for wind power modeling from meteorological wind data, Renewable and Sustainable Energy Reviews, Volume 167,2022.
Paper III	Chen, Hao; Birkelund, Yngve; Anfinsen, Stian Normann; Yuan, Fuqing	Comparative study of data-driven short-term wind power forecasting approaches for the Norwegian Arctic region. Journal of Renewable and Sustainable Energy 2021; Volum 13. (2) s.
	Chen, Hao; Birkelund, Yngve; Yuan, Fuqing	Examination of turbulence impacts on ultra-short-term wind power and speed forecasts with machine learning. Energy Reports 2021; Volum 7. Suppl. 6 s. 332-338.
Paper IV	Chen, Hao; Birkelund, Yngve; Qixia, Zhang	Data-augmented sequential deep learning for wind power forecasting. Energy Conversion and Management 2021; Volum 248. s. 1-12.
Paper V	Chen, Hao; Birkelund, Yngve	Knowledge distillation with error-correcting transfer learning for wind power prediction. (Manuscript)

Table of Contents

1 Introduction

1.1 Background

As the global economy expands at a high speed, the problems of resource shortage, energy scarcity, and environmental pollution are increasingly severe, and the global climate change issue is a growing threat to the sustainable development of human society and economy. Against this background, accelerating the transformation of energy structure to a low-carbon, clean renewable energy system is an inevitable trend [1]. Renewable energies are playing an ever-greater role in the national energy system, thanks to their sustainability, cleanliness, and improving efficiency [2]. In particular, wind and photovoltaic energy received extensive publicity and experienced rapid expansion over recent years, and are considered important forms that can complement and gradually replace the traditional power generations [3].

In particular, the Arctic is one of the regions most affected by global warming [4]. Therefore, special attention has been paid to the carbon emissions of production and transportation processes in the development of the Arctic region. In particular, the development of renewable energy sources in the region not only protects the environment but also contributes to the economic development of the region.

Several studies in recent years have shown that the Arctic is more significantly exposed to climate change than the temperate and tropical zones [5], [6], [7], [8]. A low-carbon transition in the Arctic is therefore imperative. Norway, as an economic powerhouse in the Arctic, has been at the forefront of international research and engineering practice in polar environmental energy theory in recent years. Norway's Arctic region has rich renewable energy resources, especially complementary solar and wind energy [9], [10], [11].

Wind energy, as one of the promising and technologically proven renewable energy sources, has been developing vigorously worldwide over the past decades. Wind energy has received attention from a growing number of countries for its low-cost operation and maintenance, small turbine footprint, flexibility in development scale, and rapidly decreasing electricity generation costs [12]. Global Wind Energy Council (GWEC) statistics show that the global wind power capacity is up to 743 GW by the end of 2020 and with 93 GW installations within 2020. Going forward, the wind energy industry of China, the USA, and Europe will remain continuously growing, while the industry is going to jump in South America, Africa, Latin America, and other emerging markets [13]. It is projected to account for around 12% of the global electricity supply by 2030 [14].

Efficiency and reliability are critical to making wind energy competitively possible as wind energy grows in scale [15]. One of the hallmarks of modern engineering advances is the widespread adoption of data-intensive analytical tools and applications [16]. The potential and opportunities to leverage the vast accumulated amounts of data to address efficiency and reliability challenges are being extensively discussed in wind energy research and engineering practice with the development of data science and artificial intelligence [17], [18], [19]. Technically, wind turbine technology has progressed significantly in recent years, enabling the design and deployment of larger turbines as well as the construction of wind parks in locations with cold climates and less required maintenance compared to the industry in the last century [20]. Modern wind parks have a great number of different types of sensors to keep their functioning efficient and reliable. Many anemometers, thermometers, strain sensors, and power meters are integrated into the wind turbine nacelle. Wind parks are usually also equipped with specifically constructed wind measurement towers to monitor wind conditions throughout the park. All of these monitoring tools generate vast volumes of data rapidly, providing opportunities for data science to resolve wind energy crucial issues in a new and effective way. Furthermore, the wind is a random and intermittent natural phenomenon. Therefore, the instability of wind energy is one of the most significant challenges for wind energy production and integration into the grid. Continuous innovations in data science provide comprehension of wind stochasticity and enable the design of countermeasures that could potentially yield groundbreaking progress in the wind energy industry as wind-related data continue to accumulate [21].

As mentioned above, increasing numbers of wind energy projects are being developed in colder regions. Relying more on wind energy resources in the Arctic helps achieve carbon neutrality in this cold and environmentally vulnerable area. Norway owns some of the best wind energy resources in Europe [22]. It has enormous potential for wind power generation, especially in its northern and Arctic regions. The vast sparsely vegetated lands in cold regions, the low temperature and high-density air and the excellent wind resources are favorable for the development of wind energy projects. However, developing wind energy faces many problems, such as turbine blade icing, weak local power grid, rapidly fluctuating wind speed, and direction, etc.

The Ph.D. project paid special attention to the rapidly fluctuating wind speed and direction issue in northern Norway with complex fjord topography and unique climate conditions. Norway has a cold climate and a 25,148km coastline, both of which are generally characterized by an abundance of wind energy resources, and it is with a complex terrain consisting of mountains, valleys, and fjords, making the wind change dramatically and unpredictably.

To comprehensively address the mentioned issue, under the guidance of the International Energy Agency (IEA) Wind Technology Collaboration Programme's (TCP) Research Tasks with attention on tasks as follows [23], [24]:

Task 1. Environmental and Meteorological Aspects of Wind Energy Conversion Systems (WECS) studied the environmental impact and operational safety of large-scale WECS, investigated the uncertainty in wind forecasting appropriate for day-to-day operation of WECS, and recommended design methods for selected load cases.

Task 19. Wind Energy in Cold Climates gathers and provides information about wind energy in cold climates, including project development, operation, and maintenance (O&M), health, safety, and environment (HSE), operational experiences, and recent research.

Task 36. Forecasting for Wind Energy focuses on improving the value of wind energy forecasts to the wind industry.

In terms of data science, the three items above can be organized by a formulation of wind power generation in Eq. (1).

$$f(P) = \int_x f(P \mid x)g(x)dx \qquad (1)$$

where f (.) and g (.) are probability density functions. P denotes generated wind power, and x represents wind-related factors, like wind speed, direction, etc.

The item $g(x)$ says that the understanding of distributions of wind-related factors is necessary to address the wind power distribution. The item $f(P \mid x)$ is conditional power distribution that is linked with the internal relationships between power generation and its corresponding wind factors like wind speed, air density, temperature, etc. Finally, the $f(P)$, the wind power statistic, is integrated through the x result of the formula. Since generated wind power changes with time, so, it is more precise to denote the item with a time item t as $f(P_t)$.

Specifically, the focus of this thesis is on the mainly data-driven analysis of wind turbines and site-wide data from several wind parks, mainly energy production side, in the northern Arctic of Norway, with special attention to the Fakken wind park due to data availability.

Thus, the present Ph.D. research addresses the *Data-driven Arctic wind energy analysis by statistical and machine learning approaches* based on the above three tasks from IEA TCP and Eq. 1. from data science for wind energy in three aspects:

1. Five Arctic wind parks wind resource assessment: Task 1 and Modeling $g(x)$.

2. Comprehensive wind power operation monitoring modeling for an Arctic wind park: Task 19 and Modeling $(P \mid x)$.

3. Wind power forecasting for Arctic wind park and turbines: Task 36 and Modeling P_t.

1.2 Motivation and objective

The Ph.D. work, initiated in October 2019, is mainly carried out at the Department of Technology and Safety and Arctic Centre for Sustainable Energy, UiT The Arctic University of Norway, it also benefits from the academic visits of Max Planck Institute for Meteorology and Catalonia Institute for Energy Research and academic internship of the University of Tokyo.

The motivation of this Ph.D. project is "*Whether the accurate and efficient analysis of wind energy in the Arctic, with dramatically fluctuating wind, can be achieved by developing models based on data-driven advanced statistical and machine learning methods.*"

Data-driven wind energy analysis has yielded many results [25] ,[26] ,[27], [28], [29]. But on one side, these studies are usually focused on highly populated temperate and tropical regions or areas with relatively flat topography; on the other side, wind energy studies on a particular area are compartmentalized, i.e., concerned with only one part of the full work cycle, thus leading to a lack of information on the whole spectrum of a wind park from resource evaluation to operation analysis.

Therefore, the primary objective is to develop data-driven comprehensive models for wind resource assessment, operational modeling, and power forecasting for wind parks in the Norwegian Arctic region. More specifically from the theoretical and methodological perspectives, the first is using physics and mathematical modeling to approach the question. The second is a combination and comparison of different methods in varying scenarios for deeply understanding the data characteristics and method implementations. The third is developing advanced and specialized tools with the considerations of proper data science and wind energy physics.

1.3 Problem description

The present research is a comprehensive wind energy analysis with a focus on the Norwegian Arctic conditions by advanced statistical, learning, and meteorological approaches. According to the chronological order of wind power projects, the research is conducted with attention to approaching the following problems:

Q1: Prior to wind power generation: How to assess the wind energy resources of an area in a proper approach? Task 1 and Modeling $g(x)$.

Answered by a Journal Paper I. A. ***Assessing probabilistic modeling for wind speed from numerical weather prediction model and observation in the Arctic*** and a Conference Paper I. B. ***Probability distribution for wind speed fluctuation characteristics-A case study of Northern Norway***

Q2: At present wind power is being generated: How do construct appropriate models to monitor wind power generation? Task 19 and Modeling $(P \mid x)$.

Answered by a Journal Paper II. ***Cluster-based ensemble learning for wind power modeling from meteorological wind data***

Q3: Future of wind power production: How to achieve accurate wind power forecasting to enable large-scale penetration into the grid for such green energy? Task 36 and Modeling P_t.

Answered by three Journal papers Paper III. A. ***Comparative study of data-driven short-term wind power forecasting approaches for the Norwegian Arctic region;*** Paper IV. ***Data-augmented sequential deep learning for wind power forecasting;*** Paper V. ***Knowledge distillation with error-correcting transfer learning for wind power prediction*** and a Conference papers Paper III. B. ***Examination of turbulence impacts in ultra-short-term wind power and speed forecasts with machine learning.***

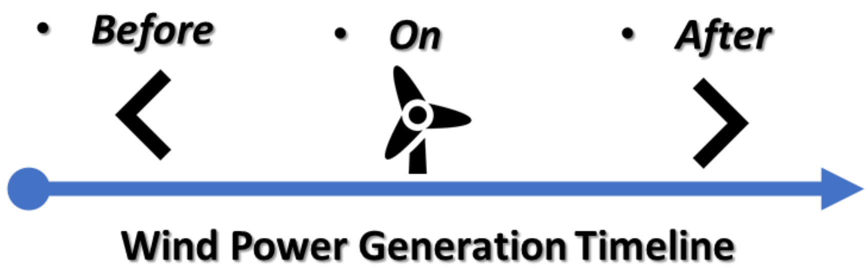


Figure 1 – Research questions timeline.

1.4 Employed technique list

As wind energy research involves highly sophisticated physical and data science, this study employs a variety of methods to address the questions asked above. They include but are not limited to Table 1.

Table 1. Summarized used techniques.

No.	Publications	Main employed techniques
Paper I	Assessing probabilistic modeling for wind speed from numerical weather prediction model and observation in the Arctic. Probability distributions for wind speed volatility characteristics: A case study of Northern Norway.	Descriptive statistics, Exploratory data analysis, Inference tests, Parametric estimation, Probability distribution.
Paper II	Cluster-based ensemble learning for wind power modeling from meteorological wind data.	Exploratory data analysis, Clustering analysis, Feature selection, Regression trees, Ensemble learning.
Paper III	Comparative study of data-driven short-term wind power forecasting approaches for the Norwegian Arctic region. Examination of turbulence impacts on ultra-short-term wind power and speed forecasts with machine learning.	Descriptive statistics, Representative machine learning algorithms, Interference tests, Univariable and multivariable regression.

Paper IV	Data-augmented sequential deep learning for wind power forecasting.	Data preprocessing, Feature engineering, Neural networks, Deep learning, Encoder-decoder networks Interference tests.
Paper V	Knowledge distillation with error-correcting transfer learning for wind power prediction.	Deep learning, Encoder-decoder networks Knowledge distillation, Error analysis, Transfer learning, Edge learning Bidirectional long short-term memory networks Inference tests.

The above techniques are cross-used and elaborated in the included papers. Due to the length limitations, the details of these techniques can be found in the attached papers when they are applied.

1.5 Target and data in brief

The major research targets are five wind parks, namely Nygårdsfjellet, Fakken, Raggovidda, Kjøllefjord, and Havøygavlen in northern Norway with the cold climate and complex terrain consisting of large mountains, valleys, and fjords. Moreover, due to the data availability, the Fakken wind park is given special attention. A descriptive overview and locations of the five sites are intuitively presented in Figure. 2 with terrain elevation around each wind park. Table 2 serves as a summarized statistical comparison in terms of their coordinates, heights installed capacity, location, and site ruggedness (RIX). The example of generated wind power data is shown in Figure 3. Northern Norway has a complex terrain consisting of fjords, mountains, and valleys that goes from the coast into a moderately high inland along the border to northern Sweden and Finland.

Nygårdsfjellet wind park is located in a valley, far from the open sea, that reached approximately 450 meters elevation. The mountains south and north of the valley limit the main wind direction to be west-east, and high wind events are expected during the winter season. Havøygavlen, Kjøllefjord, and Fakken wind parks are located close to the open sea and on relatively flat hills where large nearby fjords affect both wind direction and speed. Raggovidda wind park is also located near the open sea but on a flat mountain that does not have any vegetation.

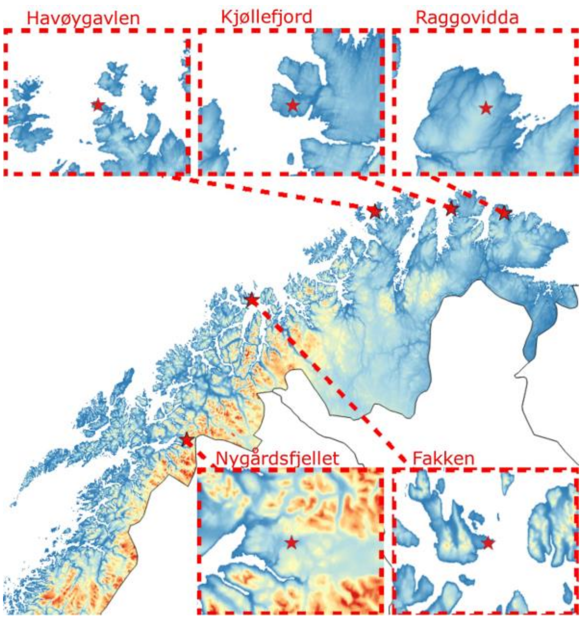


Figure 2 – The concerned five wind parks in Northern Norway. Ocean is shown in white. Color tones from blue to red show terrain heights from 0m to 2000m [30].

Table 2. Description of the concerned five wind park sites [30].

Wind Park	Location °N / °E	Height[m]	RIX	Designed power [MW]	Mean power [MW]	Standard deviation [MW]	Capacity factor of 2017
Nygårdsfjellet	68.504 / 17.879	410	0-5	32.2	11.132	11.833	34.57%
Fakken	70.098 / 20.081	95	5-10	54.0	15.239	15.858	28.22%
Raggovidda	70.769 / 29.094	440	0-5	45.0	21.782	16.869	48.40%
Kjøllefjord	70.922 / 27.268	280	10-20	39.1	12.349	12.786	31.58%
Havøygavlen	71.012 / 24.589	220	5-10	40.5	10.311	11.037	25.46%

Terrain features: Nygårdsfjellet (Inland valley, high steep mountains); Fakken (Small hill, high steep mountains and fjords); Raggovidda (Flat inland mountain, close to coast); Kjøllefjord (Low flat mountains and fjords); Havøygavlen (Flat low island, steep cliffs and fjords).

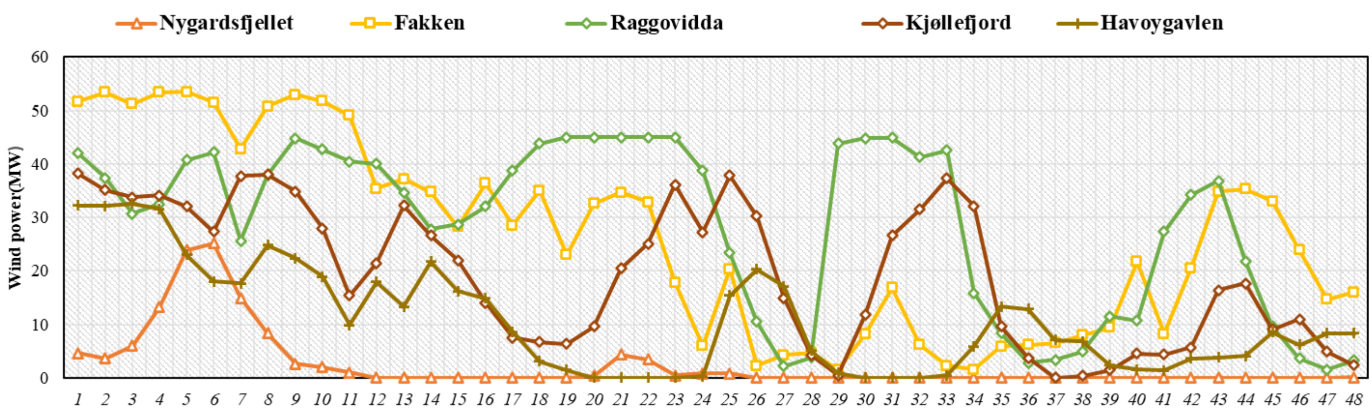


Figure 3 – Generated wind power in the first 48 hours of 2017 for the five wind parks.

The used wind data (0:00 1st January 2017 to 23:00 31st December 2017) summery is briefly shown in Table 3.

Table 3. A brief summary of used wind data.

Wind data	**Source**	**Resolution**
Wind power generation data for five wind parks	Norwegian Water Resources and Energy Directorate (NVE)	1h time resolution
Numerical weather prediction (NWP) wind data for five wind parks	MET Norway (Norwegian Meteorological Institute) MEPS (Ensemble Prediction System)	1h time resolution, 2.5 km spatial resolution
Measured wind environmental data of mast for Fakken wind park	Troms Kraft Produksjon. Local power production company, owner of Fakken wind park	10min time resolution interpolated to 1h time resolution
Measured wind environmental data of turbine for Fakken wind park	Troms Kraft Produksjon. Local power production company, owner of Fakken wind park	10min time resolution interpolated to 1h time resolution

1.6 Outline of the thesis

The remainder of the dissertation is structured as follows: Chapter 2, 3, and 4 focus on prior to wind power generation: wind resource assessment; on wind power generation: wind power curve modeling; and future wind power generation: wind power forecasting, respectively. In Chapter 5, we introduce the publications and discuss the research findings and contributions. Chapter 6 concludes the dissertation with a summary and suggestions for further work. Finally, the papers included in the dissertation can be found in the appendices.

2 Prior to wind power generation: wind resource assessment

In this chapter, Question 1: **Prior to wind power generation: How to assess the wind energy resources of wind sites in a proper approach in the Arctic?** is taken under consideration. To refine the question, it is addressed by analyzing the wind speed and its volatility distributions by statistical methods.

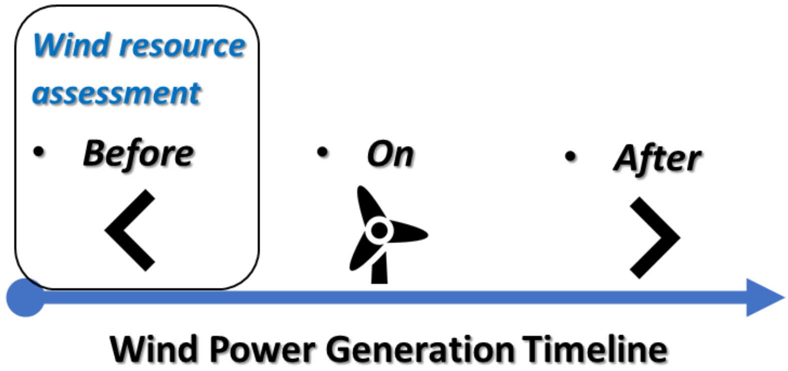


Figure 4 – Wind resource assessment.

Specifically, the question is answered by a journal paper *Assessing probabilistic modeling for wind speed from numerical weather prediction model and observation in the Arctic* (**Scientific Reports 2021, IF: 5.0**) and a conference paper *Probability distribution for wind speed fluctuation characteristics-A case study of Northern Norway* (**Elsevier Energy Reports 2021 IF: 4.9**). The main content of this chapter is drawn from the two papers mentioned above [10], [31].

2.1 Background

In Norway, multiple wind energy projects have been developed for energy markets, and many more wind parks are in the design and planning stage [32]. It is critical to developing a convincing technique for assessing the area's wind energy resources. Assessing regional wind energy potential and resources precisely is an important aspect of wind energy development since it increases investment trust in finance and risk management [33]. Because of geographic variances, wind resource potential differs from one wind park location to another. As a result, when developing durable wind power projects, a precise assessment of a wind park's wind energy potential is essential [34]. A rigorous evaluation of the potential wind speed resources of a specific location directly affects the economic value, risk assessment, turbine selection, power generation estimation of the wind park, as well as the operation and management of wind power conversion systems [35]. [10]

The probability density function (PDF) is normally applied in wind resource assessments to quantify the theoretical wind energy potential, both of which can intuitively reflect the wind speed statistical patterns [10]. Wind speed value is positive, and its PDF is skewed to the right. It should not strictly follow the normal distribution. Consequently, right-skewed ideal PDF with non-negative mean, such as Weibull or Rayleigh distributions [12], are candidates for modeling the speed.

Theoretically, a practical concept in wind engineering is named the Capacity Factor (CF). Understanding the probability distribution of wind speed is vital to computing the CF, which is computed by the mean energy production divided by the rated in a certain period, as in Eq. (2):

$$CF = \frac{P_{\text{ave}}}{P_r}, \text{where } P_{\text{ave}} = \int_0^{\infty} f(v)P(v)dv \quad (2)$$

where $f(v)$ represents the PDF of wind speed. $P(v)$ is the turbine power curve with wind speed. A wind turbine gets its maximum electricity generation when the wind speed is in the interval between the rated and cut-off speed. Knowledge of the wind speed interval is of great importance for ensuring the turbine's efficient and economical operation. So, the $f(v)$ is an important evaluation index for estimating local wind resource potential [10].

2.2 Probabilistic modeling for wind speed

- ***Assessing probabilistic modeling for wind speed from numerical weather prediction model and observation in the Arctic*** [10] see Appendix Paper I.A.

Research in a nutshell

The statistical characteristics of wind speed are essential for the practical assessment of wind energy potential and the sustainable design of wind parks. Since wind speed is variable, intermittent, and uncertain, appropriate means should be used to describe its fluctuating nature [12]. Wind is created by pressure differences between different regions, but terrain features like mountains, valleys, fjords, and other surface irregularities create disturbances, meaning that wind speeds near the ground typically fluctuate significantly. The wind speed contributing to energy production in a wind turbine surrounded by complex terrain typically changes significantly; therefore, when the time scale is short, the statistical characteristics of the wind become uncertain and difficult to predict [36]. When the time scale is long, the probabilistic distribution of wind speed is relatively stable, and the long-term statistical characteristics of wind can be determined [37]. A common way of describing the wind energy at a site is to use its annual wind speed distribution. The PDF of wind speed is vital in valuing energy production for wind power and is an important evaluation index for estimating local wind resource potential.

Most related studies have focused on PDF modeling for the observed wind speed of wind parks, and there is a lack of PDF modeling for wind speed forecasted by Numerical Weather Prediction (NWP). This is unfortunate because NWP calculations generate most of the world's wind data. Some studies have focused on using the Weibull distribution or one of three or four other similar distribution methods. However, they fail to consider the broader deployment of the PDF approach for wind speed modeling. In practice, more attention is paid to the wind speed range corresponding to the wind turbine's rated power. Despite this, few studies have applied PDF methods to analyze wind speed intervals when wind turbines are producing the maximum power, and little research has discussed wind speed distribution in the Arctic region.

In the study, we concentrate on probabilistic modeling of NWP wind speed for five wind parks in the Norwegian Arctic region and one observed wind speed, Fakken, for one of them. The results of the present study indicate that, for wind resource assessments in complex terrain, the Nakagami and Generalised extreme value distributions are recommended as the preferred models for the PDF of NWP and observed wind speed, respectively, as they showed excellent and consistent performance. In addition, the probabilistic models that reasonably describe interval wind speed differ from those of overall wind speed due to the nature of the wind: the former corresponds more to the right-side properties of the probability distribution functions.

2.3 Probability distributions for wind speed volatility

- ***Probability distributions for wind speed volatility characteristics: a case study of northern Norway*** [31] see Appendix Paper II.B.

Research in a nutshell

Due to the uncertainty and intermittency of wind, wake effects between wind turbines, and the cubic relationship between wind speed and the wind turbine-generated power, a small change in wind speed can be significantly amplified in the output wind power. The random volatility of wind is regarded as an adverse factor for wind energy [38]. This intermittency brings severe challenges to the power system's safety, power quality, and the balance of power supply and demand. Therefore, studying the volatility characteristics of wind is of great significance for improving wind power forecasting accuracy, scenario generations, and overcoming the adverse effects of wind power integration in the grid [39]. Wind speed volatility, a phenomenon that strongly affects wind power generation, has not received sufficient research attention. The typical wind energy assessment methodology lacks tools to characterize wind speed volatility on sites. The volatility analysis offers additional information about wind. The wind has different volatility characteristics at different temporal scales. Although the wind has certain seasonal and diurnal characteristics, there is no fixed volatility amplitude and cycle; its volatility has no clear rules to follow.

In the present study, we focus on statistical modeling of wind speed volatility for a wind park, Fakken, inside the Norwegian Arctic region. The probability distribution of wind volatility is overall centrally symmetrical but quite different from the normal distribution. In our cases, wind volatility is slightly left-skewed and has sharper peaks compared to the normal distribution. However, as the temporal resolution of sampling decreases, its probability distribution becomes closer to the normal distribution. Although most PDF models fail a rigorous nonparametric goodness-of-fit test based on the raw data of complex wind phenomena, the logistic and *t* distributions deliver (coefficient of determination) high R^2 and Root Mean Square Error (RMSE) approaching zero, suggesting that both distributions provide good characterizations of wind speed short-term volatility in wind energy engineering practice. Moreover, the *t* distribution has a notable advantage, and its performance is very stable with sampling time.

2.4 Summary

Question: Prior to wind power generation: How to assess the wind energy resources of wind sites in a proper approach in the Arctic?

Brief answer: The proposed comprehensive probability distribution modeling processes delivers a useful tool for assessing the wind resource, especially for wind speed and its volatility.

For wind speed assessment, we comparatively assess seven different PDFs for wind speed modeling for the five wind parks. We analyzed wind speed distributions for a wind park using NWP and observed wind data to better understand the differences in wind speed data from different resources. [10]

For wind speed volatility assessment, we use different PDFs and skewness and kurtosis moments to characterize short-term wind speed volatility at various temporal scales for Fakken. The statistical modeling of volatility assists in documenting wind's internally volatile features, especially for the wind in a cold climate and complex terrain. [31].

3 On wind power generation: wind power curve modeling

This chapter focuses on Question 2: **At present wind power is being generated: How to construct appropriate models to monitor wind power generation?** The refined question is dealt with wind power generation curve modeling from meteorological wind data with machine learning algorithms and the modeling performance is checked by statistical approaches.

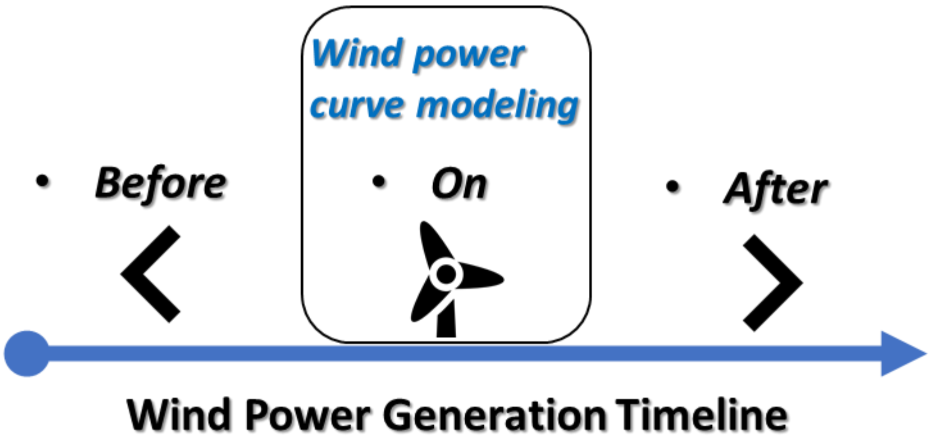


Figure 5 – Wind power curve modeling.

Specifically, the question is answered by a journal paper *Cluster-based ensemble learning for wind power modeling from meteorological wind data* (**Elsevier Renewable and Sustainable Energy Reviews 2022, IF: 16.8**). The main content of this chapter is drawn from the mentioned paper [40].

3.1 Background

Wind power is characterized by volatility, randomness, and intermittency. Establishing an accurate power model for a wind park based on the empirical mapping of weather data is important to understand the relationship between wind and wind power generation, which in turn is significant for the safe and stable operation and economic operation [41]. It is also important for having a non-parametric power curve model that can be applied as a reference profile for the online monitoring generation process [42]. In practice, for grid planning and dispatching, improving the modeling accuracy of wind power can protect the economic scheduling and power balance and can reduce the allocation of energy storage equipment capacity [43]. For wind power parks, accurate models can provide a reliable reference for power generation plans and thus improve production efficiency [26], [44]. More precise examples to describe the importance of wind power curve models are firstly the wind energy prediction can be done by forecasting wind speed initially and then converting it to predicted power by using power curve models [45], [46]. Secondly, accurate models can assist in turbine performance assessment and turbine health monitoring [47], [48]. [40]

Wind power production is a conversion from wind kinetic energy to electrical energy. Neglecting any loss in the process of conversion, the actual power output (theoretical power curve) of a wind turbine can be expressed as (3).

$$P = \begin{cases} 0 & v < v_{min} \\ \frac{1}{2} C_P \rho A v^3 & v_{min} < v < v_n \\ P_n & v_n < v < v_{max} \\ 0 & v > v_{max} \end{cases} \tag{3}$$

where P is the output power; C_P is wind energy efficiency; ρ donates the air density; A represents the effective area swept by the wind turbine blades, v is the wind speed; v_{min}, v_{max}, and v_{n} are cut-in, cut-off wind speed, and rated wind speed, respectively. P_n means the rated power for the turbine. From (3), the output of a wind turbine is mainly influenced by wind speed, air density, and swept area. Moreover, air density is primarily affected by temperature and pressure [49]. The swept area is influenced by wind direction. The functional relationship between environmental factors and power response is typically nonlinear. The nonlinearity is mainly from complex C_P that is influenced by many environmental factors [50]. Furthermore, the multiplicative relationships indicate interactions between the factors [51]. [40]

So, the complexity in Eq. (3) offers an opportunity for data science, with a data-driven nonparametric modeling approach, in the power modeling from environmental factors. From the wind park curve modeling and operation monitoring point of view, the power production can be summarized by functional relationships with the most dominant factor wind speed vector $\boldsymbol{V}$ and other weather factors $\boldsymbol{W}$, as $P=\boldsymbol{f}(\boldsymbol{V}; \boldsymbol{W})$.

3.2 Wind power modeling by cluster-based ensembles

- ***Cluster-based ensemble learning for wind power modeling from meteorological wind data*** [40] see Appendix Paper II.

Research in a nutshell

Driven by progress in computing affordability and capability and algorithmic advances, wind power can increasingly be modeled by physical, statistical, and hybrid methodologies. However, there is still room to improve these models [52]. This study presents an ensemble learning approach that combines bagging, boosting, and stacking for modeling wind power from meteorological data. To mine the inherent characteristics of the data, four clustering approaches are used to process inputs for the layered ensembles. Then, the layered cluster-based ensembles are fused within the stacking framework.

The Adaptive boosting (AdaBoost) with Random Forest (RF) model can accurately model wind power. The algorithm circumvents issues of an equal weighting of each tree in RF and

AdaBoost and allows each learner to boost incrementally, and eventually creates a model with a good generalization. The overall performance of the proposed method is proven much better than the benchmarks in the cases without clustering. As no standard methods for identifying the cluster number exist, the study uses a heuristic elbow graph, an empirical formula, and the X-means clustering to precisely determine the implied number for meteorological data.

The comparative study of Adaboost with RF-based on different clustering methods reveals, firstly, that the model with clusters significantly performs better than the model without, regardless of what clustering approach is employed. This suggests that similarities within the wind power data can correspond to similarities within the weather data. Secondly, among these clustering methods, the model with Farthest First (FF) clustering provides the best modeling results. The reason is that FF is built on finding the data point furthest from the previous centroid as the new one; in other words, it emphasizes large differences between clusters. Upon this clustering, the fluctuations among the original meteorological data are considerably diminished, which in turn corresponds to a smoother wind power output and increases the accuracy of the wind power model. The fast computability and accuracy of FF also suggest that the clustering technique can be applied to ultra-short-term wind power models. Thirdly, Canopy is the fastest among the four clustering methods and achieves comparable results. Therefore, Canopy can also serve as a favorable clustering approach when wind weather datasets are considerably large.

The wind power model is further strengthened by using stacking to fuse the layered ensembles with four clustering approaches. It can be interpreted as the two-layer stacking with four clustering methods Adaboost with RF model working as a representation learning—that is, effective features are automatically collected from raw data and fed into the second layer via multiple learners in the first layer; the second layer compiles and aggregates these features through linear regression with a regular term and effectively outputs simulations.

3.3 Summary

Question: At present wind power is being generated: How do construct appropriate models to monitor wind power generation, especially for the Arctic wind parks?

Brief answer: The wind power curve modeling for monitoring the generation can be achieved by a modeling scheme that orderly integrates three types of ensemble learning algorithms—bagging, boosting, and stacking—and clustering approaches to achieve wind power modeling from multiple wind-based meteorological factors for a wind park in the Norwegian Arctic area. The scheme involves quantifying wind turbulence, clustering meteorological data, and ensemble learning. Firstly, an effective model integrating bagging and boosting is constructed. Secondly, four prominent clustering algorithms are systematically incorporated with models to form layered cluster-based ensembles and the best clustering approach is selected. Finally, stacking is employed to fuse these ensembles with different clusters to establish a more accurate model. [40]

4 Future wind power generation: wind power forecasting

This chapter weighs heavily for the Ph.D. project, which systemically answers the Question3: **Future of wind power production: How to achieve accurate wind power forecasting?** In northern Norway, the cold climate region can, in general, be characterized by good wind resources but is challenging in wind power forecasting with the complex terrain consisting of large mountains, valleys, and fjords. So, the question is further refined into three sub-questions. The sub-questions are answered by establishing different forecasting models with machine learning and wind energy physics, and the modeling performance is checked by statistical approaches.

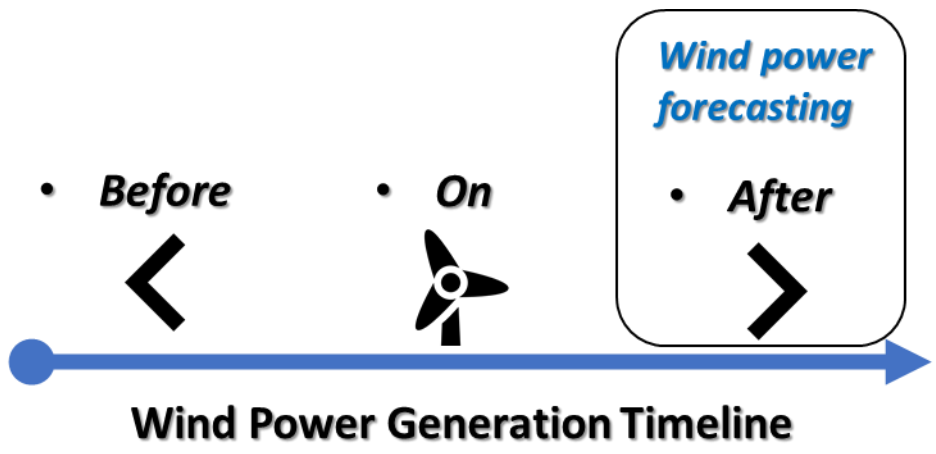


Figure 6 – Wind power forecasting.

1. Which input feature selections and existing forecasting methods are suitable for the Arctic wind scenario?

Specifically, the question is answered by a journal paper *Comparative study of data-driven short-term wind power forecasting approaches for the Norwegian Arctic region* (**AIP Journal of Renewable and Sustainable Energy 2021, IF: 2.8**) and a conference paper *Examination of turbulence impacts on ultra-short-term wind power and speed forecasts with machine learning* (**Elsevier Energy Reports 2021, IF:4.9**).

2. How to develop an advanced wind power forecasting model combining data processing and state-of-the-art deep learning algorithms?

Specifically, the question is answered by a journal paper *Data-augmented sequential deep learning for wind power forecasting* (**Elsevier Energy Conversion and Management 2021, IF: 11.5**).

3. How to quickly train an accurate but complex model and further incorporate available weather data into the model?

Specifically, the question is answered by a journal paper *Knowledge distillation with error-correcting transfer learning for wind power prediction*.

The main content of this chapter is drawn from the four papers mentioned above [53], [54], [55], and Appendix Paper V.

4.1 Background

Along with the electricity grid adding wind power penetrations, the unstable grid factors also increase, which are undesirable to the power system's effective and safe operations. So, it is crucial to use proper methods to understand wind power production and harness proper methods to make forecasts of the electricity generated by wind parks. However, due to the inherent characteristics of random nonlinear fluctuations in wind speed variation, it is difficult to obtain satisfactory prediction performance. To integrate wind energy into the power supply system efficiently and economically, it is vital to accurately make the wind power prediction.

Wind power prediction can be divided into ultra-short-term prediction, short-term prediction, medium-term prediction, and long-term prediction [56]. Ultra-short-term forecasts are predictions made from a few minutes to 30 minutes in advance; the short-term are forecasts made from 30 minutes to 48 hours ahead, the medium-term refers to predictions made days, weeks, or months earlier, and the long-term is made years in advance. Wind power forecasting methodology is generally divided into physical, statistical, and hybrid approaches. [57] The first predicts wind power by extensive numerical computation of physical equations. It is based on fluid dynamics and uses Numerical Weather Prediction (NWP) data such as wind speed and pressure, and geoinformation like ground roughness and altitude. The method performs best in medium or long-term forecasting and applies to the wind resource assessment of new wind parks that lack historical observations. The statistical approach aims to establish linear or nonlinear patterns within wind data sequences that can be utilized in forecasting. In particular, machine learning-based wind power forecasting methods developed in recent years are widely applied. The hybrid approach is a combination of the former categories and has shown its edge profoundly [58].

Regarding data science, wind power forecasting could be simplified as a multivariable regression project, in which wind power time series are autoregressed, while wind speed and other weather factors serve as the supplementing information to the autoregression. Updating the related weather forecasts from NWP of the predicted time is a crucial feature in the forecasting based on an extensively cited reference by Giebel and Kariniotakis [59].

Generally, the fundamental multistep forecasting model *f*(.) with timestep $i+n$ is expressed as:

$$\hat{P}_{i+n} = f\left(P_{i-j}; W_{i-j}; NWP_{i+n}\right) + \varepsilon_n \tag{4}$$

where i means the base current time and with each i, j is the previous time steps. $\hat{P}_{i+n}$ is n timestep ahead forecasting power, W is the measured wind, NWP donates wind data computed by the mesoscale NWP model. ε_n is the model error. [53], [55]

4.2 Comparative investigation of wind power forecasting

The section firstly applies various machine learning methods for wind power forecasting and comprehensively compares the performance of models categorized by whether consider weather factors. Then, it investigates the use of turbulence intensity for the predictions of wind power and speed for a wind park in the Arctic.

4.2.1 Comparative study of data-driven forecasting approaches

- ***Comparative study of data-driven short-term wind power forecasting approaches for the Norwegian Arctic region*** [53] see Appendix Paper III. A.

Research in a nutshell

This study conducts a systemic comparative study on univariate and multivariate wind power forecasting for five wind parks inside the Arctic area.

For the univariate time series wind power prediction in these cases, the Persistence Model (PE) approach and machine learning methods do not have a considerable difference in performance. The Support Vector Regression (SVR) and MultiLayer Perceptron (MLP) function equally well with the PE model. The machine learning algorithms that perform best in Mean Average Error (MAE) are SVR and SVR optimized with Genetic Algorithm (GA-SVR), whose average MAE is almost the same as for the PE model. The machine learning algorithm that performs best in Root Mean Square Error (RMSE) is MLP. SVR, Radial Basis Functions (RBF), GA-SVR, and Long Short-Term Memory neural networks (LSTM) also have lower RMSE than the PE model has. This generally means the PE model has more large errors in the prediction procedure. Our result also validates the conclusion from research [60] in the wind engineering field. The conclusion is learning algorithms do not deliver on their promise for univariate time series prediction, and the classical statistical methods even perform better. The phenomena may be explained that for univariate series, the complex methods often overlearn the training set and create overfitting in the testing.

For the multivariate wind power forecasting in our cases, the model considers methodological or topographic factors by taking the mesoscale NWP wind data as inputs. Compared to the corresponding algorithm in the univariate case, the multivariate model has a lower MAE and results in a smaller RMSE. When the predictive time increases, the multivariate models are more stable than the PE model, especially in the metric of RMSE. These prove that the multivariate model entirely exceeds the PE model and the univariate model. Furthermore, the sophisticated ensemble and deep learning algorithm demonstrate their superiority in dealing with complex and multivariate pattern recognitions in complicated wind power forecasting problems.

The NWP wind data are generated with mesoscale (2.5km), which is larger than the area of our wind parks. However, adding this local weather information can still obviously optimize the performance of forecasting models. The improvements in the penetration of NWP data in wind power prediction can be explained from two aspects: firstly, from the Bayesian theory [61], the introduced NWP wind information can provide a priori probability information to make more precise wind power (corresponding to posterior probability) predictions; secondly, NWP wind data can be regarded as the simulating wind conditions of the whole wind park, which add useful information in the predictive process.

4.2.2 Turbulence impacts on the forecasting

- ***Examination of turbulence impacts on ultra-short-term wind power and speed forecasts with machine learning*** [54] see Appendix 4.2.

Research in a nutshell

In the present study, we focus on various machine learning autoregressive approaches to realize forecasts for wind power and speed for a wind park inside the Norwegian Arctic regions. The performances of different machine learning algorithms in predicting ultra-short-term wind power and speed are satisfactory but not significantly different in general. 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 in the forecast model, given that it is a subsidiary factor and increases computational burdens.

Since the wind park understudy has a complex topography, there may be turbulence interactions, both natural and generated by the wind turbines. As a whole wind park, these turbulent currents could cancel each other out. It is advantageous to examine the 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.

4.3 An advanced wind power forecasting framework

- ***Data-augmented sequential deep learning for wind power forecasting*** [55] see Appendix Paper IV.

Research in a nutshell

The present study returns to the physical process of wind power generation, the statistical characteristics of wind data, and the nature of deep learning to approach the forecasting problem. After synthesizing numerous data augmentation methodologies and drawing on multiple state-of-the-art advances in sequential data prediction, the robust and efficacious encoder-decoder deep neural networks with stacking LSTM units are proposed for wind turbine power forecasting in the Arctic. It initially scrutinizes the usefulness of data augmentation approaches in wind power forecasting and proposes a multi-input and multi-output prediction algorithm with verified superiority in multistep forecasting five wind turbines with various topologies.

The proposed seq2seq-based deep Encoder-Decoder Long Short-Term Memory neural networks (EDLSTM) enable highly effective and robust multistep power forecasting, by highlighting the sequential dependence of the problem, for wind turbines under different terrain conditions.

Since EDLSTM is a complex deep learning model, its strength requires so-called big data. It is demonstrated that five-fold expansions of the primary data with data augmentations statistically boost neural network-based NN (Three-layer backpropagation Neural Networks), LSTM, BA (Bionic optimized neural networks constructed Adaboost), and EDLSTM wind power forecasting capabilities. The boost is particularly evident in EDLSTM, where, on average, the performance of the data-augmented model provides better forecasting with lower RMSE, which is 10.2% smaller than its counterpart without data augmentation. This boosting can be interpreted as expanding the training set, it is equivalent to adding a regular term to the loss function when training models, which can effectively avoid overfitting. Besides, due to the stochasticity involved in data augmentations, the learned model built on the techniques presents better robustness. Moreover, the data-augmented EDLSTM edges over the benchmarks since the proposed EDLSTM further learned deeper information, like signal decompositions, of the wind data by mentioned augmentation techniques.

The impact of the eight data augmentation approaches employed, three physics-oriented and five data-oriented, on wind power prediction is forecasting arithmetic sensitive. For the proposed well-performing EDLSTM, various augmentations can approximately, by over 12%, boost the forecasting qualification rate at the 90% threshold. But augmentations improve the forecasting performance to slightly different degrees when evaluated by RMSE and MAE, and generally, data-oriented augmentations outperform physics-oriented ones. Among data-oriented augmentations, the results illustrate that EDLSTM's forecasting RMSE is significantly decreased even by simply appending noisy and randomly perturbing or moving data the same way as sophisticated statistical data decomposition and learning data generation, however, as per MAE, the latter two provide overall closer predictions to the real power.

4.4 An integrated but fast wind power forecasting framework

- ***Knowledge distillation with error-correcting transfer learning for wind power prediction*** see Appendix Paper. V

Research in a nutshell

The presented study fully addresses the above concerns by innovatively developing a framework for predicting wind turbine power with mathematical derivation. It is the first time, through Knowledge Distillation (KD), to deploy a large sequential deep learning forecasting model with multiple inputs and outputs for big data in wind parks on a fast-running small-scale turbine forecasting model; the framework fully exploits the prediction error value and ingeniously downscales the NWP information corresponding to wind parks non-explicitly to the turbine scale by transfer learning and primitive and inverse function transformation. In other words, A deep learning wind power prediction framework bridges large (park with big data) and small-scale (turbine with small data) forecasting through a proposed KD regression approach, and maps park-scale weather forecasts non-explicitly to turbine-scale by Transfer Learning (TL)-based prediction error corrections. The effectiveness and quickness of the framework are experimentally verified on wind turbines in different terrains. Furthermore, the framework has extensive applicability in other fields since it does not involve specific energy-physics and geographical factors. The following conclusions are drawn from experiments on turbine predictions under five types of topography in a park in the Arctic.

The proposed multistep power prediction framework is extraordinarily effective and robust by leveraging the data and reinforcing the nonlinear capabilities of the model based on the experimental comparisons. Compared to its competitors, the overall average effectiveness, in RMSE, is respectively improved from high to low: 23.9%, 21.1%, 14.9%, 7.9%, and 3.3%. The effectiveness is also verified with a metric with physical meaning, Qualification Rate at the 90% threshold (QR90). Moreover, KD-TL, thanks to its adequate utilization of weather forecast information, yields satisfactory outcomes in predicting wind turbines in complex terrain, which is normally challenging as well. Finally, the complexity and response time of KD-TL decreases multiplicatively over its closest competitor, which enables the proposed approach to have extensive strengths in engineering deployment.

4.5 Summary

Question: Future of wind power production: How to achieve accurate wind power forecasting, especially for the Arctic wind parks?

Brief answer: Traditional machine learning methods could be to achieve short-term wind power forecasts for the Arctic region, in which these forecasts could be improved to some extent by applying appropriate wind meteorological data to the forecasting models. Expanding the amount of data with appropriate data processing techniques and loading the data into large deep learning models allows for more accurate wind power predictions in different terrain conditions. With a variety of transfer learning techniques, it is possible to refine the raw data information better and transfer large accurate but slow training forecasting models to smaller and faster ones for realizing rapid and efficient wind power forecasting.

5 Research findings and contribution

Throughout this chapter, a summary of the finished academic papers of the Ph.D. project and their research findings and contributions, by presenting abstracts and contribution parts of these papers.

5.1 List of publications

The Ph.D. thesis includes **5** scientific journal articles that are finished or published through international scientific journals on the related topics. Except for Paper III.A which is published in a subscribing way, whose postprint version is self-archived at the Norwegian academic repository Cristin. The others are fully open access.

Additionally, some results are also presented and published **7** short communication journal articles. 2021 & 2022 International Conferences on Technologies and Materials for Renewable Energy, Environment and Sustainability (TMREES), 2021 International Conference on Power and Energy Systems Engineering (CPESE), 2021 & 2022 International Conference on Electrical Engineering and Green Energy (CEEGE), 2022 International Conference on Clean Energy and Electrical Systems (CEES), 2021 Asia Conference on Automation Engineering (ACAE)

In total, the author of this Ph.D. thesis successfully published **11** peer-reviewed scientific papers and a finished manuscript during his Ph.D. period.

The present dissertation includes the following five scientific journal and two conference papers that served as supplementary materials for two of the five journal papers:

Paper I.

A. *Chen, Hao; Birkelund, Yngve; Anfinsen, Stian Normann; Staupe-Delgado, Reidar; Yuan, Fuqing*. Assessing probabilistic modeling for wind speed from numerical weather prediction model and observation in the Arctic. *Scientific Reports 2021; Volum 11. (7613) s.*

B. *Chen, Hao; Anfinsen, Stian Normann; Birkelund, Yngve; Yuan, Fuqing. Probability distributions for wind speed volatility characteristics: A case study of Northern Norway. Energy Reports 2021; Volum 7. s. 248-255. CONFERENCE*

Paper II.

Chen, Hao. Cluster-based ensemble learning for wind power modeling from meteorological wind data, Renewable and Sustainable Energy Reviews, Volume 167,2022.

Paper III.

A. *Chen, Hao; Birkelund, Yngve; Anfinsen, Stian Normann; Yuan, Fuqing.* *<u>Comparative study of data-driven short-term wind power forecasting approaches for the Norwegian Arctic region.</u>* *Journal of Renewable and Sustainable Energy 2021; Volum 13.(2) s.*

B. *Chen, Hao; Birkelund, Yngve; Yuan, Fuqing.* *<u>Examination of turbulence impacts on ultra-short-term wind power and speed forecasts with machine learning</u>. Energy Reports 2021; Volum 7. Suppl. 6 s. 332-338. CONFERENCE*

Paper IV

Chen, Hao; Birkelund, Yngve; Qixia, Zhang. *<u>Data-augmented sequential deep learning for wind power forecasting.</u>* *Energy Conversion and Management 2021; Volum 248. s. 1-12.*

Paper V

Chen, Hao; Birkelund, Yngve. *<u>Knowledge distillation with error-correcting transfer learning for wind power prediction.</u>*

Table 4. Contribution of the Ph.D. candidate to the papers in the Ph.D. thesis.

Topic	**Paper I A, B**	**Paper II**	**Paper III A, B**	**Paper IV**	**Paper V**
Conceptualization	√	√	√	√	√
Data curation				√	√
Formal analysis	√	√	√	√	√
Investigation	√	√	√	√	√
Methodology	√	√	√	√	√
Resources		√		√	√
Software	√	√	√	√	√
Visualization	√	√	√	√	√
Validation	√	√	√	√	√
Paper writing	√	√	√	√	√
Submitting and revising	√	√	√	√	√

5.2 Research findings and contribution

This part presents research summaries and findings in the form of abstracts and/or contributions from papers I through V that are included in this dissertation.

- **Prior to wind power generation: Assessing the wind energy resources of the Norwegian Arctic wind parks**

Paper I. A. Assessing probabilistic modeling for wind speed from numerical weather prediction model and observation in the Arctic [10]

In this research, different probability density functions are used to model wind speed for five wind parks in the Norwegian Arctic region. A comparison between wind speed data from numerical weather prediction models and measurements is made, and probability analysis for the wind speed interval corresponding to the rated power, which is largely absent in the existing literature, is presented. The results of the present study suggest that no single probability function outperforms across all scenarios. However, some differences emerged from the models when applied to different wind parks. The Nakagami and Generalised extreme value distributions were chosen for the numerical weather predicted prediction and the observed wind speed modeling, respectively, due to their superiority and stability compared with other methods. This paper, therefore, provides a novel direction for understanding the numerical weather prediction wind model and shows that its speed statistical features are better captured than those of real wind.

The main contributions of this paper can be summarized as follows:

1. This study is the first to conduct a PDF modeling analysis of wind speed intervals associated with wind turbine rated power, with a particular focus on differences between interval and overall wind speed modeling.

2. This paper compares wind speed distributions based on wind data from NWP and measurements. Wind speed distributions provide an intriguing and well-established approach to analyzing wind speed resources, and this paper investigates their use for NWP models in the context of complex coastal terrain.

3. The present study can assist in a more detailed understanding of PDF applications in wind speed modeling, as seven ideal distributions are used to model wind speed. Moreover, it offers an insight into the potential for renewable energy utilization in the Arctic by conducting natural resource modeling in this area, with clear implications for practice, policy, and future project implementation.

Paper I. B. Probability distributions for wind speed volatility characteristics: a case study of northern Norway [31]

Wind speed volatility, a phenomenon that strongly affects wind power generation, has not received sufficient research attention. In this paper, a framework for studying short-term wind speed volatility with statistical analysis and probabilistic modeling is constructed for an existing wind park in Northern Norway. It is found that unlike the characteristics of wind power volatility, wind speed volatility cannot be described by the normal distribution. The reason is that even though the probability distribution of wind speed volatility is centrally symmetric, it is much more centrally concentrated and has thicker tails. After comparing three distributions corresponding to different sampling periods, this paper suggests utilizing the t distribution, with average modeling RMSE less than 0.006 and (coefficient of determination) R^2 exceeding 0.995 and with the best modeling scenario of temporal resolution, the 30 mins has an RMSE of 0.0051 and an R^2 of 0.997, to explore the fluctuating characteristics of wind speed more accurately and effectively. The statistical modeling of volatility assists in documenting wind's internally volatile features, especially for the wind in a cold climate and complex terrain.

- **At present wind power is being generated: Constructing an appropriate model to monitor wind power generation**

Paper II. Cluster-based ensemble learning for wind power modeling with meteorological wind data [40]

Based on the idea that similar wind conditions lead to similar wind powers; this paper constructs a modeling scheme that orderly integrates three types of ensemble learning algorithms, bagging, boosting, and stacking, and clustering approaches to achieve wind power modeling from multiple wind-based meteorological factors in a wind farm. The paper also investigates the applications of different clustering algorithms and methodologies to determine cluster numbers in the modeling. The results reveal that all ensemble models with clustering exploit the intrinsic information in wind data and thus outperform models without clustering by approximately 15% on average in modeling wind power. The model with the best-performing Farthest First clustering is computationally rapid and with an improvement of around 30% compared with the baselines. Given the diversity introduced by clustering algorithms, the power modeling performance is further boosted by about 5% by introducing stacking that fuses ensembles with varying clusters. The proposed modeling framework thus demonstrates promise by delivering efficient and robust performance on the targeted problem.

The principal contributions of this paper are thus as follows.

1. This paper experimentally proves that farthest first clustering is a distinctive approach in clustering wind data for power modeling compared to K-means, expectation-maximization, and Canopy clustering algorithms. It shows that even the worst-performing layered cluster-based ensemble outperforms the one without clustering. This indicates the similarities and dissimilarities in wind data. However, even though these data are not related to an individual wind turbine, they can still be significantly reflected in wind power in an implicit form.

2. Given the differences in results of different clustering algorithms, the paper proposes fusing layered ensembles with varying clusters with two-layer stacking to formalize a model that exceeds the optimal single clustering method. The stacking can more efficiently and quickly

address the complex mapping task of nonlinear relationships between meteorological wind data and wind power.

3. The paper builds a procedure for determining the cluster number with a heuristic elbow chart, an empirical formula, and an X-means clustering approach. The procedure may be further developed and refined into a technique for identifying cluster numbers on other problems.

4. AdaBoost boosting with random forest bagging as its weak learner is apposite in the wind power models. These tree-based algorithms are computationally fast and parameter insensitive compared to the network-based ones. The proposed AdaBoost model statistically outperforms linear, neural network, and benchmark Adaboost approaches.

5. The quantization of wind turbulence intensities—both wind speed and direction that are rarely considered in related research—is applied to wind power modeling in a novel manner. The study finds that both intensities can serve as new features for considering wind volatility in the modeling.

- **Future of wind power production: Developments of accurate wind power forecasts for wind parks and turbines**

Paper III. A. Comparative study of data-driven short-term wind power forecasting approaches for the Norwegian Arctic region [53]

This work conducts a systemic comparative study on univariate and multivariate wind power forecasting for five wind parks inside the Arctic area. The development of wind power in the Arctic can help reduce greenhouse gas emissions in this environmentally fragile region. In practice, wind power forecasting is essential to maintain the grid balance and optimize electricity generation. This study firstly applies various learning methods for wind power forecasting. It comprehensively compares the performance of models categorized by whether considering weather factors in the Arctic. Nine different representative types of machine learning algorithms make several univariate time series forecasting, and their performance is evaluated. It is demonstrated that machine learning approaches have an insignificant advantage over the persistence method in the univariate situation. With numerical weather prediction wind data and wind power data as inputs, the multivariate forecasting models are established and made one hour to six hours in advance predictions. The multivariate models, especially with the advanced learning algorithms, show their edge over the univariate model based on the same algorithm. Although weather data are mesoscale, they can contribute to improving wind power forecasting accuracy. Moreover, these results are generally valid for the five wind parks, proving the models' effectiveness and universality in this regional wind power utilization. Additionally, there is no clear evidence that predictive model performance is related to wind parks' topographic complexity.

The main contributions of this study can be summarized as follows:

1. For brief experimental univariate power forecasts. The persistence model and nine machine learning benchmarking algorithms are researched in forecasting models and compared to their performance from an algorithm perspective. We find the persistence model performs almost equally to machine learning models in our cases. The result also proves conclusions from Ref.[60]; those classical methods may dominate univariate time series forecasting. However, we find that its performance drops more quickly with the forecast time step rises. Considering the contingency of parameters tuning and computational complexity of the learning algorithm, it is suggested that statistical modeling methods should be primarily considered in forecasting.

2. The multivariate models with mesoscale NWP wind data, although the data resolution scale is larger than the wind park area, as inputs can slightly gain prediction accuracy compared with the univariate models with the same algorithm. Moreover, the multivariate models reduce performance slower than the univariate models, which indicates the informative complementary role that the weather data play in the model.

3. The five Arctic wind parks have different complex terrains and climates. The wind park with complex terrains implies that the NWP wind results are not as accurate as their counterparts in plain landscapes. However, there is no significant evidence that prediction results are related to the ruggedness index of wind parks from our results.

Paper III. B. Examination of turbulence impacts on ultra-short-term wind power and speed forecasts with machine learning [54]

Wind turbines' economic and secure operation can be optimized through accurate ultra-short-term wind power and speed forecasts. Turbulence, considered 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 park 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, especially for the Arctic regions.

Paper IV. Data-augmented sequential deep learning for wind power forecasting [55]

With excellent automatic pattern recognition and nonlinear mapping ability for big data, deep learning is increasingly employed in wind power forecasting. However, salient realities are that in-situ measured wind data are relatively expensive and inaccessible and correlation between steps is omitted in most multistep wind power forecasts. This paper is the first time that data augmentation is applied to wind power forecasting by systematically summarizing and proposing both physics-oriented and data-oriented time-series wind data augmentation approaches to considerably enlarge primary datasets and develops deep encoder-decoder long short-term memory networks that enable sequential input and sequential output for wind power forecasting. The proposed augmentation techniques and forecasting algorithm are deployed on five turbines with diverse topographies in an Arctic wind park, and the outcomes are evaluated against benchmark models and different augmentations. The main findings reveal that on one side, the average improvement in RMSE of the proposed forecasting model over the benchmarks is 33.89%, 10.60%, 7.12%, and 4.27% before data augmentations, and increases to 40.63%, 17.67%, 11.74%, and 7.06%, respectively, after augmentations. The other side unveils that the effect of data augmentations on prediction is intricately varying, but for the proposed model with and without augmentations, all augmentation approaches boost the model outperformance from 7.87% to 13.36% in RMSE, 5.24% to 8.97% in MAE, and similarly over 12% in QR90. Finally, data-oriented augmentations, in general, is slightly better than physics-driven ones.

The principal contributions of the present study paper are as follows:

1. We exhaustively develop a seq2seq deep learning predictive end-to-end model with inputs of historical wind speed and power data and wind speed from NWP as well as simultaneously interrelated outputs of multistep, futuristic wind power. The model is based on an encoder-decoder constructed with LSTM and shows its superiority in forecasting power.

2. It is demonstrated that the impact of various augmentation approaches is different in each forecasting algorithm. Augmentations somewhat increase linear, like persistence model errors. Nonetheless, augmentations improve the performance, most notably the proposed deep learning model, of neural networks-based algorithms, where data-oriented augmentations generally contribute greater than physics-oriented ones.

3. The data augmentations combined with the proposed and benchmark forecasting models are utilized to predict power generated by five turbines in various landscapes. The results are analyzed by rigorous statistical methods and indicate that the augmentations and the proposed forecasting model have wind engineering values and potentially extensive applicability in other energy fields.

Paper V. Knowledge distillation with error-correcting transfer learning for wind power prediction

Wind power prediction, especially for turbines, is vital for the operation, controllability, and economy of electricity companies. Hybrid methodologies combining advanced data science with weather forecasting have been incrementally applied to the predictions. Nevertheless, individually modeling massive turbines from scratch and downscaling weather forecasts to turbine size are neither easy nor economical. Aiming at it, this work proposes a novel framework with mathematical underpinnings for turbine power prediction. It is the first time to incorporate knowledge distillation into energy forecasting, enabling accurate and economical constructions of turbine models by learning knowledge from the well-established park model. Besides, park-scale weather forecasts non-explicitly are mapped to turbines by transfer learning of predicted power errors, achieving model correction for better performance. The proposed framework is deployed on five turbines featuring various terrains in an Arctic wind park, the results are evaluated against the competitors of ablation investigation. The major findings reveal that the proposed framework yields performance boosts from 3.3 % to 23.9 % over its competitors. This advantage also exists in terms of wind energy physics and computing efficiency, which are verified by the prediction quality rate and calculation time.

The work is further continuous research for Paper IV. As Paper IV yielded a delicate forecasting structure promising multistep prediction. However, there are still the following points where enhancements are possible according to our further investigations.

1. Deep learning-based models require multiple layers of uniquely designed neural network structures to realize the intrinsic features of the data [62]. And training large deep networks is very time-consuming and computationally intensive. Therefore, it is proposed that certain pre-training techniques, such as Knowledge Distillation [63], can be adopted to *distill* useful information, knowledge, from the large pre-trained, whole park, *teacher* model and *condense* it into smaller scale turbine prediction, *student*, models.

2. The NWP information in hybrid prediction models is usually for the whole wind park and does not precisely reflect the individual turbine's future meteorological condition. Besides, the wake effect (wind is strongly perturbed, decreased kinetic energy and added turbulence behind blades of a turbine), together with the turbulence induced by the micro-scale topography in complex terrain, thereby rendering further forecasting difficulties. Typically, single wind turbine meteorological modeling considers NWP results and simulates turbine wind conditions with Computational Fluid Dynamics [64], [65]. This paper will bypass this complex approach based on multiple physical assumptions and indirectly integrate the meteorological information of turbines into the whole prediction model by data science.

3. Wind power forecasting is essentially reducible to a regression problem, so regression diagnostics in statistics, especially error analysis and correction, could be incorporated into the prediction model. Hence, this paper achieves the detection and forecasting of prediction errors through advanced deep learning approaches. Moreover, weather information is ingeniously embedded into the final prediction model by the errors correcting.

6 Conclusion and future work

6.1 Concluding remarks

As increasing numbers of wind power projects are developed in the northern Norway Arctic and large amounts of wind power operational data are being accumulated, data-driven comprehensive analysis of wind power in this complex terrain and cold climate region becomes potentially feasible. The present Ph.D. thesis is motivated by the challenge "Whether the accurate and efficient analysis of wind energy in the Arctic, with dramatically fluctuating wind, can be achieved by developing models based on data-driven advanced statistical and machine learning methods?" and guided by IEA TCP and a proposed wind energy theoretical equation from data science to answer three research questions of prior to, at present, and in future wind power generation for wind parks in northern Norway.

Firstly, for question 1: Prior to wind power generation: How to assess the wind energy resources of wind sites in a proper approach in the Arctic? Its concluded answer: The proposed comprehensive probability distribution modeling processes delivers a useful tool for assessing the wind resource, especially for wind speed and its volatility both for the wind data from measurements and numerical modeling.

Then, for question 2: At present wind power is being generated: How to construct appropriate models to monitor wind power generation for the Arctic wind park? Its concluded answer: The proposed wind power curve scheme involves quantifying wind turbulence, clustering meteorological data, and ensemble learning. The scheme orderly integrates three types of ensemble learning algorithms—bagging, boosting, and stacking—and clustering approaches to achieve accurate wind power modeling from multiple wind-based meteorological factors for a wind park in the Norwegian Arctic area.

Furthermore, for question 3: Future of wind power production: How to achieve accurate wind power forecasting for the Arctic wind park? Its concluded answer: Traditional machine learning methods could be to achieve short-term wind power forecasts for the Arctic region, which these forecasts could be improved to some extent by applying appropriate wind meteorological data to the forecasting models. Expanding the amount of data with appropriate data processing techniques and loading the data into large deep learning models allows for more accurate wind power predictions in different terrain conditions. With a variety of transfer learning techniques, it is possible to refine the raw data information better and transfer large accurate but slow training forecasting models to smaller and faster ones for realizing rapid and efficient wind power forecasting.

These three research questions, arranged in the wind power generation timeline, have been thoroughly and in-depth investigated through the present thesis and attached papers. The project achieved the target goal of developing data-driven Arctic wind energy analysis by statistical and machine learning approaches for wind parks, especially Fakken, in the Norwegian Arctic.

6.2 Future work

There are still some potential research perspectives on data-driven approaches for Arctic wind energy. The following are a few recommendations and comments for future investigations:

Firstly, for wind resources assessment, the statistical properties of wind speed in Arctic wind sites are well investigated. However, wind resources are also affected by other environmental factors such as wind direction and air density. The development of a joint distribution of environmental variables based on more advanced multivariate statistical techniques could provide a more comprehensive picture of wind resources in the region.

Then, for wind power modeling, future research is further improving the proposed wind power curve model with a deep optimization for base learners of ensemble learning architecture and their combination algorithms to deliver faster and more accurate modeling. Another direction is to incorporate an in-the-now power modeling approach and meteorological data with historical wind power to achieve efficacious very short-term power forecasting for the Arctic wind park.

Moreover, the existing studies covered the targeted Arctic wind park and wind turbine power prediction. Further research could be: 1. exploring the mathematical properties within complex deep learning forecasting algorithms and giving confidence intervals for further developing probabilistic forecasts; 2. investigating how to better exploit the potentials of weather forecast data in wind power forecasting; 3. incorporating data science, weather forecasting and satellite observations of weather and other multi-source models is also a good direction.

In terms of wind energy engineering practice, the possibility of integrating data-driven wind resource assessment, wind power curve modeling, and wind power forecasting into an open-access or commercial user-friendly platform, to maximize resource utilization is also worth further consideration.

References

[1] F. Kern and K. S. Rogge, "The pace of governed energy transitions: Agency, international dynamics and the global Paris agreement accelerating decarbonisation processes?," *Energy Research & Social Science,* vol. 22, pp. 13-17, 2016.

[2] R. U. Ayres and E. H. Ayres, *Crossing the energy divide: moving from fossil fuel dependence to a clean-energy future*. Pearson Prentice Hall, 2009.

[3] D. Infield and L. Freris, *Renewable energy in power systems*. John Wiley & Sons, 2020.

[4] N. Wunderling, M. Willeit, J. F. Donges, and R. Winkelmann, "Global warming due to loss of large ice masses and Arctic summer sea ice," *Nature communications,* vol. 11, no. 1, pp. 1-8, 2020.

[5] J. E. Box *et al.*, "Key indicators of Arctic climate change: 1971–2017," *Environmental Research Letters,* vol. 14, no. 4, p. 045010, 2019.

[6] A. Witze, "The Arctic is burning like never before--and that's bad news for climate change," *Nature,* vol. 585, no. 7825, pp. 336-338, 2020.

[7] L. Sun, M. Alexander, and C. Deser, "Evolution of the global coupled climate response to Arctic sea ice loss during 1990–2090 and its contribution to climate change," *Journal of Climate,* vol. 31, no. 19, pp. 7823-7843, 2018.

[8] W. F. Vincent, "Arctic climate change: Local impacts, global consequences, and policy implications," in *The Palgrave handbook of Arctic policy and politics*: Springer, 2020, pp. 507-526.

[9] B. Babar, R. Graversen, and T. Boström, "Solar radiation estimation at high latitudes: Assessment of the CMSAF databases, ASR and ERA5," *Solar Energy,* vol. 182, pp. 397-411, 2019.

[10] H. Chen, Y. Birkelund, S. N. Anfinsen, R. Staupe-Delgado, and F. Yuan, "Assessing probabilistic modelling for wind speed from numerical weather prediction model and observation in the Arctic," *Scientific Reports,* vol. 11, no. 1, pp. 1-11, 2021.

[11] K. Solbakken, B. Babar, and T. Boström, "Correlation of wind and solar power in high-latitude arctic areas in Northern Norway and Svalbard," *Renewable Energy and Environmental Sustainability,* vol. 1, p. 42, 2016.

[12] P. Jain, *Wind energy engineering*. McGraw-Hill Education, 2016.

[13] M. DeCastro *et al.*, "Europe, China and the United States: Three different approaches to the development of offshore wind energy," *Renewable and Sustainable Energy Reviews,* vol. 109, pp. 55-70, 2019.

[14] W. Dong, H. Sun, J. Tan, Z. Li, J. Zhang, and Y. Y. Zhao, "Short-term regional wind power forecasting for small datasets with input data correction, hybrid neural network, and error analysis," *Energy Reports,* vol. 7, pp. 7675-7692, 2021.

[15] T. Stehly, P. Beiter, and P. Duffy, "2019 cost of wind energy review," National Renewable Energy Lab.(NREL), Golden, CO (United States), 2020.

[16] D. P. Mandic *et al.*, "Data fusion for modern engineering applications: An overview," in *International Conference on Artificial Neural Networks*, 2005: Springer, pp. 715-721.

[17] G. J. Herbert, S. Iniyan, E. Sreevalsan, and S. Rajapandian, "A review of wind energy technologies," *Renewable and sustainable energy Reviews,* vol. 11, no. 6, pp. 1117-1145, 2007.

[18] F. Dincer, "The analysis on wind energy electricity generation status, potential and policies in the world," *Renewable and sustainable energy reviews,* vol. 15, no. 9, pp. 5135-5142, 2011.

[19] M. Sawant, S. Thakare, A. P. Rao, A. E. Feijóo-Lorenzo, and N. D. Bokde, "A review on state-of-the-art reviews in wind-turbine-and wind-farm-related topics," *Energies,* vol. 14, no. 8, p. 2041, 2021.

[20] G. W. E. C. GWEC, "Global offshore wind report 2020," *URL https://gwec. net/global-offshore-wind-report-2020,* 2020.
[21] Y. Ding, *Data science for wind energy*. CRC Press, 2019.
[22] B. Blindheim, "Implementation of wind power in the Norwegian market; the reason why some of the best wind resources in Europe were not utilised by 2010," *Energy policy,* vol. 58, pp. 337-346, 2013.
[23] E. Peltola, H. Holttinen, S. Rissanen, and C. Murphy-Levesque, "IEA Wind TCP 2016 Overview," in *IEA Wind TCP 2016 Annual Report*, 2017, pp. 6-27.
[24] K. L. Dykes *et al.*, "IEA wind TCP: Results of IEA wind TCP workshop on a grand vision for wind energy technology," National Renewable Energy Lab.(NREL), Golden, CO (United States), 2019.
[25] A. Kusiak and Z. Zhang, "Short-horizon prediction of wind power: A data-driven approach," *IEEE Transactions on Energy Conversion,* vol. 25, no. 4, pp. 1112-1122, 2010.
[26] E. T. Renani, M. F. M. Elias, and N. A. Rahim, "Using data-driven approach for wind power prediction: A comparative study," *Energy Conversion and Management,* vol. 118, pp. 193-203, 2016.
[27] H. Long, L. Wang, Z. Zhang, Z. Song, and J. Xu, "Data-driven wind turbine power generation performance monitoring," *IEEE Transactions on Industrial Electronics,* vol. 62, no. 10, pp. 6627-6635, 2015.
[28] D. Zhang, L. Qian, B. Mao, C. Huang, B. Huang, and Y. Si, "A data-driven design for fault detection of wind turbines using random forests and XGboost," *Ieee Access,* vol. 6, pp. 21020-21031, 2018.
[29] D. Astolfi, F. Castellani, A. Lombardi, and L. Terzi, "Data-driven wind turbine aging models," *Electric Power Systems Research,* vol. 201, p. 107495, 2021.
[30] Y. Birkelund, S. Alessandrini, Ø. Byrkjedal, and L. D. Monache, "Wind power predictions in complex terrain using analog ensembles," 2018.
[31] H. Chen, S. N. Anfinsen, Y. Birkelund, and F. Yuan, "Probability distributions for wind speed volatility characteristics: A case study of Northern Norway," *Energy Reports,* vol. 7, pp. 248-255, 2021.
[32] A. Duffy *et al.*, "Land-based wind energy cost trends in Germany, Denmark, Ireland, Norway, Sweden and the United States," *Applied Energy,* vol. 277, p. 114777, 2020.
[33] E. W. E. Association, *Wind energy-the facts: a guide to the technology, economics and future of wind power*. Routledge, 2012.
[34] J. Yuan, "Wind energy in China: Estimating the potential," *Nature Energy,* vol. 1, no. 7, pp. 1-2, 2016.
[35] J. Wang, J. Hu, and K. Ma, "Wind speed probability distribution estimation and wind energy assessment," *Renewable and sustainable energy Reviews,* vol. 60, pp. 881-899, 2016.
[36] M. Bilal, Y. Birkelund, M. Homola, and M. S. Virk, "Wind over complex terrain–Microscale modelling with two types of mesoscale winds at Nygårdsfjell," *Renewable Energy,* vol. 99, pp. 647-653, 2016.
[37] B. Safari and J. Gasore, "A statistical investigation of wind characteristics and wind energy potential based on the Weibull and Rayleigh models in Rwanda," *Renewable Energy,* vol. 35, no. 12, pp. 2874-2880, 2010.
[38] T. Yang, "Optimal sizing of the hybrid energy storage system aiming at improving the penetration of wind power," in *2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, 2016: IEEE, pp. 2358-2362.
[39] S. Z. Moghaddam, "Generation and transmission expansion planning with high penetration of wind farms considering spatial distribution of wind speed," *International Journal of Electrical Power & Energy Systems,* vol. 106, pp. 232-241, 2019.

[40] H. Chen, "Cluster-based ensemble learning for wind power modeling from meteorological wind data," *Renewable and Sustainable Energy Reviews,* vol. 167, p. 112652, 2022.

[41] Z. Tian, Y. Ren, and G. Wang, "Short-term wind power prediction based on empirical mode decomposition and improved extreme learning machine," *Journal of Electrical Engineering & Technology,* vol. 13, no. 5, pp. 1841-1851, 2018.

[42] A. Marvuglia and A. Messineo, "Monitoring of wind farms' power curves using machine learning techniques," *Applied Energy,* vol. 98, pp. 574-583, 2012.

[43] W.-Y. Chang, "A literature review of wind forecasting methods," *Journal of Power and Energy Engineering,* vol. 2, no. 04, p. 161, 2014.

[44] M. Ferreira, A. Santos, and P. Lucio, "Short-term forecast of wind speed through mathematical models," *Energy Reports,* vol. 5, pp. 1172-1184, 2019.

[45] S. S. Soman, H. Zareipour, O. Malik, and P. Mandal, "A review of wind power and wind speed forecasting methods with different time horizons," in *North American power symposium 2010*, 2010: IEEE, pp. 1-8.

[46] Q. Zhou, C. Wang, and G. Zhang, "Hybrid forecasting system based on an optimal model selection strategy for different wind speed forecasting problems," *Applied Energy,* vol. 250, pp. 1559-1580, 2019.

[47] Y. He and A. Kusiak, "Performance assessment of wind turbines: data-derived quantitative metrics," *IEEE Transactions on Sustainable Energy,* vol. 9, no. 1, pp. 65-73, 2017.

[48] E. Lapira, D. Brisset, H. D. Ardakani, D. Siegel, and J. Lee, "Wind turbine performance assessment using multi-regime modeling approach," *Renewable Energy,* vol. 45, pp. 86-95, 2012.

[49] C. Jung and D. Schindler, "The role of air density in wind energy assessment–A case study from Germany," *Energy,* vol. 171, pp. 385-392, 2019.

[50] A. W. Manyonge, R. Ochieng, F. Onyango, and J. Shichikha, "Mathematical modelling of wind turbine in a wind energy conversion system: Power coefficient analysis," 2012.

[51] G. Ofualagba and E. Ubeku, "Wind energy conversion system-wind turbine modeling," in *2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century*, 2008: IEEE, pp. 1-8.

[52] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renewable energy,* vol. 37, no. 1, pp. 1-8, 2012.

[53] H. Chen, Y. Birkelund, S. N. Anfinsen, and F. Yuan, "Comparative study of data-driven short-term wind power forecasting approaches for the Norwegian Arctic region," *Journal of Renewable and Sustainable Energy,* vol. 13, no. 2, p. 023314, 2021.

[54] H. Chen, Y. Birkelund, and F. Yuan, "Examination of turbulence impacts on ultra-short-term wind power and speed forecasts with machine learning," *Energy Reports,* vol. 7, pp. 332-338, 2021.

[55] H. Chen, Y. Birkelund, and Q. Zhang, "Data-augmented sequential deep learning for wind power forecasting," *Energy Conversion and Management,* vol. 248, p. 114790, 2021.

[56] J. Jung and R. P. Broadwater, "Current status and future advances for wind speed and power forecasting," *Renewable and Sustainable Energy Reviews,* vol. 31, pp. 762-777, 2014.

[57] H. Liu, C. Chen, X. Lv, X. Wu, and M. Liu, "Deterministic wind energy forecasting: A review of intelligent predictors and auxiliary methods," *Energy Conversion and Management,* vol. 195, pp. 328-345, 2019.

[58] S. Hanifi, X. Liu, Z. Lin, and S. Lotfian, "A critical review of wind power forecasting methods—past, present and future," *Energies,* vol. 13, no. 15, p. 3764, 2020.

[59] G. Giebel and G. Kariniotakis, "Wind power forecasting—A review of the state of the art," *Renewable energy forecasting,* pp. 59-109, 2017.

[60] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and Machine Learning forecasting methods: Concerns and ways forward," *PloS one,* vol. 13, no. 3, p. e0194889, 2018.
[61] J. M. Bernardo and A. F. Smith, *Bayesian theory*. John Wiley & Sons, 2009.
[62] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
[63] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531,* vol. 2, no. 7, 2015.
[64] Y. Wang, Y. Liu, L. Li, D. Infield, and S. Han, "Short-term wind power forecasting based on clustering pre-calculated CFD method," *Energies,* vol. 11, no. 4, p. 854, 2018.
[65] L. Liu and Y. Liang, "Wind power forecast optimization by integration of CFD and Kalman filtering," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects,* vol. 43, no. 15, pp. 1880-1896, 2021.

Appendix

No.	Publications
Paper I	A. Assessing probabilistic modeling for wind speed from numerical weather prediction model and observation in the Arctic. Scientific Reports 2021; Volum 11. (7613) s.
	B. Probability distributions for wind speed volatility characteristics: A case study of Northern Norway. Energy Reports 2021; Volum 7. s. 248-255.
Paper II	Cluster-based ensemble learning for wind power modeling from meteorological wind data, Renewable and Sustainable Energy Reviews, Volume 167,2022.
Paper III	A. Comparative study of data-driven short-term wind power forecasting approaches for the Norwegian Arctic region. Journal of Renewable and Sustainable Energy 2021; Volum 13. (2) s.
	B. Examination of turbulence impacts on ultra-short-term wind power and speed forecasts with machine learning. Energy Reports 2021; Volum 7. Suppl. 6 s. 332-338.
Paper IV	Data-augmented sequential deep learning for wind power forecasting. Energy Conversion and Management 2021; Volum 248. s. 1-12.
Paper V	Knowledge distillation with error-correcting transfer learning for wind power prediction.

Paper I. A.

Assessing probabilistic modeling for wind speed from numerical weather prediction model and observation in the Arctic

https://doi.org/10.1038/s41598-021-87299-4

Assessing probabilistic modelling for wind speed from numerical weather prediction model and observation in the Arctic (Postprint)

Hao Chen*,[1], Yngve Birkelund[1], Stian Normann Anfinsen[2], Reidar Staupe-Delgado[1], and Fuqing Yuan[1]

[1]*Department of Technology and Safety,* [2]*Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø 9019, Norway*

Abstract

Mapping Arctic renewable energy resources, particularly wind, is important to ensure the transition into renewable energy in this environmentally vulnerable region. The statistical characterisation of wind is critical for effectively assessing energy potential and planning wind park sites and is, therefore, an important input for wind power policymaking. In this article, different probability density functions are used to model wind speed for five wind parks in the Norwegian Arctic region. A comparison between wind speed data from numerical weather prediction models and measurements is made, and a probability analysis for the wind speed interval corresponding to the rated power, which is largely absent in the existing literature, is presented. The results of the present study suggest that no single probability function outperforms across all scenarios. However, some differences emerged from the models when applied to different wind parks. The Nakagami and Generalised extreme value distributions were chosen for the numerical weather predicted prediction and the observed wind speed modelling, respectively, due to their superiority and stability compared with other methods. This paper, therefore, provides a novel direction for understanding the numerical weather prediction wind model and shows that its speed statistical features are better captured than those of real wind.

Keywords: resource modelling, wind energy, numerical weather prediction, probability distribution, Arctic

1.Introduction

With the growing reliance on renewable energy resources in many regions of the world, studying the predictability of renewable energy is becoming progressively important.[1] As one of the cleanest renewable energy sources, wind energy has attracted growing attention worldwide.[2] In Norway, multiple wind energy projects have been developed for energy markets, and many more wind parks are in the design and planning stage. It is, therefore, essential to create a compelling and effective method for evaluating wind energy resources in the region. Accurately assessing local wind energy potential and resources is a crucial part of wind energy development and enhances investor confidence in financial feasibility and risk acceptability [3] Wind resource potential varies considerably from one wind park site to another due to geographical and topographical differences. Therefore, an accurate assessment of a wind park's wind energy potential is necessary when developing sustainable wind power projects.[4] A rigorous evaluation of the potential wind speed resources of a specific location directly affects the economic value, risk assessment, turbine selection, power generation estimation of the wind park, as well as the operation and management of wind power conversion systems.[5] In other words, proper attention to site selection is crucial for long-term sustainability gains in wind power investments, in addition to social priorities due to the recognised nuisance conflicts that have previously arisen in the context of wind power developments.

Since wind speed is variable, intermittent and uncertain, appropriate means should be used to describe its fluctuating nature.[6] The probability density function (PDF) and the related cumulative distribution function (CDF) are often used in wind resource assessments to quantify the theoretical wind energy potential of an area. Both of them intuitively reflect the statistical characteristics of wind speed. Wind is created by pressure differences between different regions, but terrain features like mountains, valleys, fjords and other surface irregularities create disturbances, meaning that wind speeds near the ground typically fluctuate significantly. The wind speed contributing to energy production in a wind turbine surrounded by complex terrain typically changes significantly; therefore, when the time scale is short, the statistical characteristics of the wind become uncertain and difficult to predict.[7] When the time scale is long, the probabilistic distribution of wind speed is relatively stable, and the long-term statistical characteristics of wind can be determined.[8] A common way of describing the wind energy at a site is to use its annual wind speed distribution. The PDF of wind speed is vital in valuing energy production for wind power and is an important evaluation index for estimating local wind resource potential.

1.1. Related work

Some prior research on wind resources is based on probability distribution methods for specific regions with varying wind conditions and wind power potential. The two-parameter Weibull distribution is a widely used statistical distribution in wind engineering;[9, 10] however, the fitting results are not optimal for some regions, which results in a substantial difference between the estimated annual power generation and the actual yearly power generation.[5] This suggests that distribution may not be a good representation of some wind conditions or some sites. Elsewhere, researchers have expressed concerns over the role of case studies for practical wind engineering purposes. Aries, N. et al. conducted a case study of four sites' wind speed with eight distribution models for four sites in Algeria and found that the Generalised extreme value and Gamma Distributions were the most reliable base on the root mean square error evaluation.[11] Wang, J. et al. compared parametric and nonparametric models for wind speed probability distribution by taking four sites in central China as examples and showed the edge of nonparametric models.[5] Alavi, O. et al. demonstrated that the most suitable probability distributions for evaluating wind speed were not the same based on five different measurement stations distributed in the east and south-east of Iran.[12] Ayodele, T. et al. used the Weibull distribution to estimate the wind resource in a coastal area of South Africa with complex terrain.[13] Gualtieri, G. et al. focused on coastal locations in Southern Italy and used the Weibull distribution extrapolating model to assess wind resource to the turbine hub height.[14] Allouhi, A. et al. also used the Weibull distribution to describe the frequencies of actual wind data in six coastal locations in Morocco based on hourly wind speeds and directions data of five years between 2011 and 2015.[15] Jiménez, P et al. found that atmospheric stability plays a major role in controlling the shape of the wind speed distribution. The authors showed that the shape wind speed measured from a combination of long-term wind observations and numerical simulations is strongly modulated by the numerical atmospheric scales.[16]

Most studies in this field have focused on PDF modelling for the observed wind speed of wind parks, and there is a lack of PDF modelling for wind speed forecasted by Numerical Weather Prediction (NWP). This is unfortunate because NWP calculations generate the vast majority of the world's wind data. Some studies have focused on using the Weibull distribution or one of three or four other similar distribution methods. However, they fail to consider the broader deployment of the PDF approach for wind speed modelling. In practice, more attention is paid to the wind speed range corresponding to the wind turbine's rated power. Despite this, few studies have applied PDF methods to analyse wind speed intervals when wind turbines are producing the maximum power, and little research has discussed wind speed distribution in the Arctic region.

1.2. Objectives

In this research, we comparatively assess seven different probability distributions for wind speed modelling, some of which are classical, while others have rarely been used to estimate the wind speed distribution for five wind parks in the Norwegian Arctic coastal region. To improve the understanding of differences in wind speed data from different resources, we compared wind speed distributions for a wind park with NWP and observed wind data.

The main contributions of this paper can be summarised as follows:

1. The present study is the first to conduct a PDF modelling analysis of wind speed intervals associated with wind turbine trunnion rated power, with a particular focus on differences between interval and overall wind speed modelling.
2. This paper compares wind speed distributions based on wind data from NWP and measurements. Wind speed distributions provide an intriguing and well-established approach to analyse wind speed resources, and this paper investigates their use for NWP models in the context of complex coastal terrain.
3. The present study can assist in a more detailed understanding of PDF applications in wind speed modelling, as seven ideal distributions are used to model wind speed. Moreover, it offers an insight into the potential for renewable energy utilisation in the Arctic by conducting natural resource modelling in this area, with clear implications for practice, policy and future project implementation.

The paper is organised as follows: In Section 2, we describe the wind data and their sites to provide the context of the study. In Section 3, we elaborate on the methodological aspects of the study, while Section 4 outlines the experimental process. Section 5 presents the results and main implications of the study and reflects on their relevance for research and practice. The final and concluding section summarises the most important elements of the research.

2. Description of wind park and wind speed data

In the present study, we focus on five wind parks in the Norwegian Arctic regions. The second and third columns of Table 1 list their locations and the site ruggedness index (RIX).[17] The RIX is an empirical parameter for measuring the complexity of nearby terrain and is typically used in fluid modelling or in numerical weather models to indicate identify turbulence may interfere with the model results. In our case, this was based on a fraction of the area within a 2 km radius around the location with a more than 30% degree inclination and was extracted from a Norwegian mapping of wind resources.

2.1. Numerical Weather Prediction

Scandinavian meteorological institutes use an operational numerical weather prediction **(**NWP) forecast known as the Meteorological Ensemble Prediction System (MEPS). The NWP model is a complex mathematical model of the atmosphere that divides the Earth's surface into grids.[18] The grid's spatial resolution determines how meteorological processes are simulated with different accuracy levels, which limits the quality of the forecasts. A study conducted by the Norwegian Meteorological Institute demonstrated that the higher-resolution regional NWP model did not lead to better wind power forecasts for some Norwegian wind parks.[19] Therefore, in the present study, we considered NWP data with a horizontal resolution of 2.5 km as a relatively coarse resolution in wind predictions.

2.2. Data description

NWP wind data from the five wind parks were extracted from Norwegian Meteorological Institute's operational MEPS models. The predictions initiated at 00, 06, 12 and 18 UTC and were made available for operational use about 2 hours later. The observed wind data were offered by Troms Kraft AS – the power company that operates Fakken wind park. In the present study, we combined the forecast data into a single time series with hourly wind speed data from 0:00 on 1 January 2017 to 23:00 on 31 December 2017. The year is with wind conditions of northern Norway coastline are not significantly different from the previous fifteen years. Table 1 provides a summary of the overall data. The coefficient of variation is defined as the standard deviation divided by the mean.

Table 1. The location of wind parks and statistics of their wind speed

Wind park	**Location °N /°E**	**RIX**	**Mean (m/s)**	**Standard Deviation (m/s)**	**Min (m/s)**	**Max (m/s)**	**Coefficient of Variation**	**Skewness**	**Kurtosis**
Nygårdsfjellet	68.504/ 17.879	0 to 5	8.096	5.038	0.032	31.481	0.622	0.775	3.815
Raggovidda	70.098/ 20.081	5 to 10	9.490	5.101	0.107	32.430	0.538	0.666	3.361
Kjøllefjord	70.769/ 29.094	0 to 5	7.900	4.213	0.080	25.508	0.533	0.704	3.453
Havøygavlen	70.922/ 27.268	10 to 20	8.335	4.434	0.097	26.926	0.532	0.709	3.359
Fakken (NWP)	71.012/ 24.589	5 to 10	6.948	3.885	0.097	33.686	0.559	1.164	5.960

Fakken (MEASURE)	71.012/ 24.589	5 to 10	7.687	4.627	0.000	35.100	0.602	1.338	5.660

3. Methodology

3.1. Wind energy

In wind engineering, the capacity factor (CF) is particularly useful when conducting a fast evaluation at the early design and planning stages of a wind park. Understanding the probability distribution of wind speed is essential to calculating the CF of wind parks. The CF is calculated from the average value of wind energy produced divided by the rated wind power by a wind turbine in a certain period, which may be read from the following equations (1), (2), (3):

$$CF = \frac{P_{\text{ave}}}{P_r} \quad (1)$$

$$P_{\text{ave}} = \int_0^\infty f\,(v) P(v) dv \quad (2)$$

$$P(v) = \begin{cases} P_r & v_r < v \le v_o \\ P_r \times g(v) & v_i < v \le v_r \\ 0 & v \le v_i, v > v_0 \end{cases} \quad (3)$$

where $f(v)$ is the PDF of wind speed, which is the main target of this research $P(v)$ reflects the turbine power curve used to describe the power fluctuations related to wind speed. v_i, v_r, v_o, and P_r represent the cut-in speed, the rated speed, the cut-off speed, and the rated power, respectively. [5, 20] The $g(v)$ is a multiplier increasing from 0 to 1 within the interval, that depends on the wind turbine specification. A wind turbine reaches its maximum power output when the wind speed is in the interval between the rated and cut-off speed. Adequate knowledge of the wind speed interval corresponding to the wind turbine's rated power is important for ensuring the efficient and economical operation of the turbine. Therefore, aside from the wind speed distribution modelling, we also paid special attention to wind speed in this rated power interval.

3.2. Probability distribution

Tables 2 and 3 offer brief mathematical expressions of the seven ideal probability distributions used in the present study. These distributions are defined as follows:

- Gamma distribution is a two-parameter continuous probability distribution.[21]
- Generalised extreme value distribution (GEV) is a continuous probability distribution developed with extreme value theory. [22]
- Nakagami distribution is a generalised two parameters probability distribution model proposed by Nakagami Minoru.[23] It has received extensive attention, as it can model a broad range of fading channel conditions and describe many empirical data sets.[24]
- Normal distribution, also called Gaussian distribution, is a continuous probability distribution for ideally describing a real-valued random variable.[25]

- Rayleigh distribution essentially describes the distribution of the mode of a stochastic two-dimensional vector when the two components of the vector are independent, have the same variance and are normally distributed with zero means.[26, 27]
- T distribution is commonly used to estimate the mean of a small population that is normally distributed, where the standard deviation is unknown.[28]
- Weibull distribution is the theoretical basis for reliability analysis and life inspections and is widely used for describing the probability distribution of wind speed.[29]

Table 2. The mathematical expressions of distributions

Distribution	**PDF**	**Note**	**CDF**
Gamma	$f(x;a,b) = \frac{1}{b^a \Gamma(a)} \int_0^x t^{a-1} e^{-\frac{t}{b}} dt$ where $t(x) = \begin{cases} \left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi} & \xi \neq 0 \\ e^{-(x-\mu)/\sigma} & \xi = 0 \end{cases}$	a is a shape parameter b is a scale parameter and $\Gamma(.)$ is the Gamma function	$F(x) = e^{-t(x)}$
GEV	$f(x;\mu,\sigma,\xi) = \frac{1}{\sigma} t(x)^{\xi+1} e^{-t(x)}$	μ is a location parameter $\sigma > 0$ is a scale parameter ξ is a shape parameter	$F(x;a,b) = \frac{1}{b^a \Gamma(a)} \int_0^x t^{a-1} e^{-\frac{t}{b}} dt$
Nakagami	$f(x;m,\Omega) = \frac{2m^m}{\Gamma(m)\Omega^m} x^{2m-1} exp\left(-\frac{m}{\Omega} x^2\right), \forall x \geq 0$	$m \geq 1/2$ is a shape parameter $\Omega \geq 0$is a spread parameter	$F(x;m,\Omega) = \frac{\gamma\left(m, \frac{m}{\Omega} x^2\right)}{\Gamma(m)}$ where $\gamma(.)$ is the Incomplete Gamma function and $\Gamma(.)$ is the Gamma function
Normal	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	μ is the mean $\sigma$is the standard division erf(.) is the error function	$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{2}\left[1 + erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$
Rayleigh	$f(x;\sigma) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}, \quad x \geq 0$	$\sigma > 0$ is a scale parameter	$F(x;\sigma) = 1 - e^{-x^2/(2\sigma^2)}$
t	$f(x;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\nu\pi}} \frac{1}{\left(1+\frac{x^2}{\nu}\right)^{\frac{\nu+1}{2}}}$	$\nu > 0$ is the number of degrees of freedom and $\Gamma(.)$ is the Gamma function	$F(x;\nu) = \int_{-\infty}^x \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\nu\pi}} \frac{1}{\left(1+\frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}} dt$

Weibull	$f(x;\lambda,k)=\begin{cases}\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x\geq 0\\ 0 & x<0\end{cases}$	$k\ >0$is a shape parameter and $\lambda>0$ is a scale parameter	$F(x;\lambda,k)=\begin{cases}1-e^{-(x/\lambda)^k} & x\geq 0\\ 0 & x<0\end{cases}$

Table 3. The mean and variance expressions of distributions

Distribution	**Mean**	**Variance**
Gamma	ab	ab^2
GEV	$\begin{cases}\mu+\sigma(g_1-1)/\xi & if\,\xi\neq 0,\xi<1\\ \mu+\sigma\gamma & if\,\xi=0\\ \infty & if\,\xi\geq 1\end{cases}$ where $g_k=\Gamma(1\ -\ k\xi)$, and γ is Euler's constant	$\begin{cases}\sigma^2(g_2-g_1^2)/\xi^2 & if\,\xi\neq 0,\xi<\frac{1}{2}\\ \sigma^2\frac{\pi^2}{6} & if\,\xi=0\\ \infty & if\,\xi\geq\frac{1}{2}\end{cases}$
Nakagami	$\frac{\Gamma\left(m+\frac{1}{2}\right)}{\Gamma(m)}\left(\frac{\Omega}{m}\right)^{1/2}$	$\Omega\left(1-\frac{1}{m}\left(\frac{\Gamma\left(m+\frac{1}{2}\right)}{\Gamma(m)}\right)^2\right)$
Normal	μ	σ
Rayleigh	$\sigma\sqrt{\frac{\pi}{2}}$	$\frac{4-\pi}{2}\sigma^2$
t	0, for $\nu>1$	$\frac{\nu}{\nu-2}$, for $\nu>2$
Weibull	$\lambda\Gamma(1+\frac{1}{k})$	$\lambda^2\left[\Gamma\left(1+\frac{2}{k}\right)-\left(\Gamma\left(1+\frac{1}{k}\right)\right)^2\right]$

3.3. Parametric estimation

Parametric estimation for the PDFs of wind speed refers to the assumption that a specific probability distribution model can describe the wind speed, where the parameters of the model are estimated based on available wind speed data. Several parametric estimation methods can be used in wind engineering, including the moment method, empirical approach, graphical method and maximum likelihood method.[30] In a study comparing six methods for estimating Weibull parameters to fit wind data, the maximum likelihood method was, on the whole, shown to provide more accurate estimations than other methods in tests with both simulated and observation data.[31] Therefore, we used the maximum likelihood method to identify the parameters for all seven probability density functions in the present study. The Maximum likelihood estimation (MLE) method can be explained as follows:[32] If $\{X_1, X_2, \cdots, X_n\}$ is an independent and identically distributed sample from a population with PDF $f(x|\theta_1,\cdots\theta_k)$. The likelihood function is defined by equation (4):

$$L(\theta\mid X)=L(\theta_1,\cdots\theta_k\mid x_1\cdots x_n)=\prod_{i=1}^{n} f\,(x_i\mid\theta_1,\cdots\theta_k)\quad (4)$$

If $L(\theta \mid X)$ is differentiable in θ, then the values of θ_i that minimize $L(\theta_1, \cdots \theta_k \mid x_1 \cdots x_n)$ are solutions of possible candidates θ_i for the MLE are calculated by equation (5):

$$\frac{\partial}{\partial \theta_i} L(\theta \mid X) = 0, \quad i = 1,2, \cdots k \quad (5)$$

3.4. Performance comparison

The Friedman test is used to check for differences in performance across multiple trials to accurately compare the modelling performance of different probability distributions in different wind parks.[33] In particular, column effects are checked after adjusting with possible row effects. The significant level of the Friedman test was set as 0.01 in the present study.

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

H_a: The column data have a significant difference.

The statistic F is shown as in (6):

$$F = \frac{12n}{k(k+1)} \left[\sum_{i=1}^{k} r_i^2 - \frac{k(k+1)^2}{4} \right] \quad (6)$$

where k is the number of columns, r_i is the mean value of row i. It follows $\chi^2_{(k-1)}$ under H_o.

4. Experiment setup and evaluation

4.1. Estimation of PDF

We used 0.5 m/s as the bin size to create histograms of hourly wind speed throughout the whole year for the wind speed data from the NWP models at the five wind parks and the observed wind speed of the Fakken wind park. The MLE method was then used to estimate the parameters required to define each of the theoretical PDFs, as described in Section 3.3. A one-sample Kolmogorov–Smirnov test (K-S test) was conducted to confirm whether the original wind speed data came from calculated ideal distributions by comparing the CDF of the original data and fitted ideal distributions.

Since the histogram is discrete, the kernel distribution is typically taken an empirical nonparametric PDF modelling method based on the original data. It harnesses the kernel functions (typically Gaussian function) to connect adjacent bins of the histogram to create continuous PDFs of the data. Unlike histograms, the kernel distribution approximates infinitesimal length sampling, thereby reducing sampling errors between each bin. Therefore, it can be considered a more real historical distribution of raw data. Graphically, we named this 'PDF smoothing', and it was defined by a kernel function $K(\cdot)$ and a bandwidth d in (7):

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

We conducted two separate modelling analyses – overall and interval wind speed PDF fitting – to achieve a better understanding of the probabilistic characteristic of wind. Based on five wind park features and wind turbine power curve characteristics of our six cases, we choose the wind speed interval related to the rated power, with the rated speed of 10 m/s and the cut-off speed of 20 m/s, which are typical parameters for commercial medium-size wind turbines.

4.2. Performance evaluation criteria

The K–S test is a nonparametric statistical test based on cumulative distribution function that tests whether a distribution is different from a type of ideal distribution.[34] A nonparametric test is used to test a hypothesis.[35]

H_o: $\{X_1, X_2, \cdots, X_n\}$ has a given continuous distribution.

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

The K–S test is constructed from the statistic in equation (8):

$$D = \sup_x |F_0(x) - F(x)| \quad (8)$$

where $F_0(x)$ represents CDF of the given ideal distribution, and $F(x)$ is CDF of $\{X_1, X_2, \cdots, X_n\}$. The test statistic is compared to critical values from the theoretical distribution of the Brownian bridge (If a Brownian motion, which is the random motion of particles suspended in a medium, starts at a certain point and returns to the starting point at the end, the process is called Brownian bridge.).[36]

To evaluate and compare the different examined performance of PDFs for modelling the wind speed, the mean absolute error (MAE) and root mean square error (RMSE) were used to calculate the probability density difference between parametric ideal distributions and the original PDF smoothing with speed unit of 0.01 m/s. Both are negatively oriented metrics, indicating that the smaller values are related to better performance. The MAE and RMSE determine the accuracy of a model by calculating averages of the absolute and square difference between the histogram-based PDF from the NWP and observed data and different theoretical PDF models, as expressed in equations (9) and (10). The RMSE assigns a higher weight to larger errors due to the square calculation, which penalises more significant model errors and indicates whether the model has a significant error variance.[37] Hence, the MAE and RMSE provide a comprehensive representation of a model's performance.

$$MAE = \frac{\sum_{i=1}^{n} |modeling_i - smoothing_i|}{n} \quad (9)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (modeling_i - smoothing_i)^2}{n}} \quad (10)$$

5. Results and Discussion

5.1. PDF modelling graph

Histograms and PDFs graphs are shown in Fig. 1 to show the estimated ideal PDFs with the MLE method for different cases. Discontinuous histograms are represented by bar charts (for clarity, we ignored the kernel distribution curves in these figures), and the fitted probability distribution curves are shown in different colours. As can be seen, although the wind speed distributions of different wind parks varied, they had some similarities. It is also clear that different probabilistic models provide differing fits to wind speeds. In particular, when comparing (e) and (f), the actual wind speed is more centrally concentrated and possesses a thicker tail. Due to the scarcity of data, this phenomenon could only be considered empirical for the Fakken site.

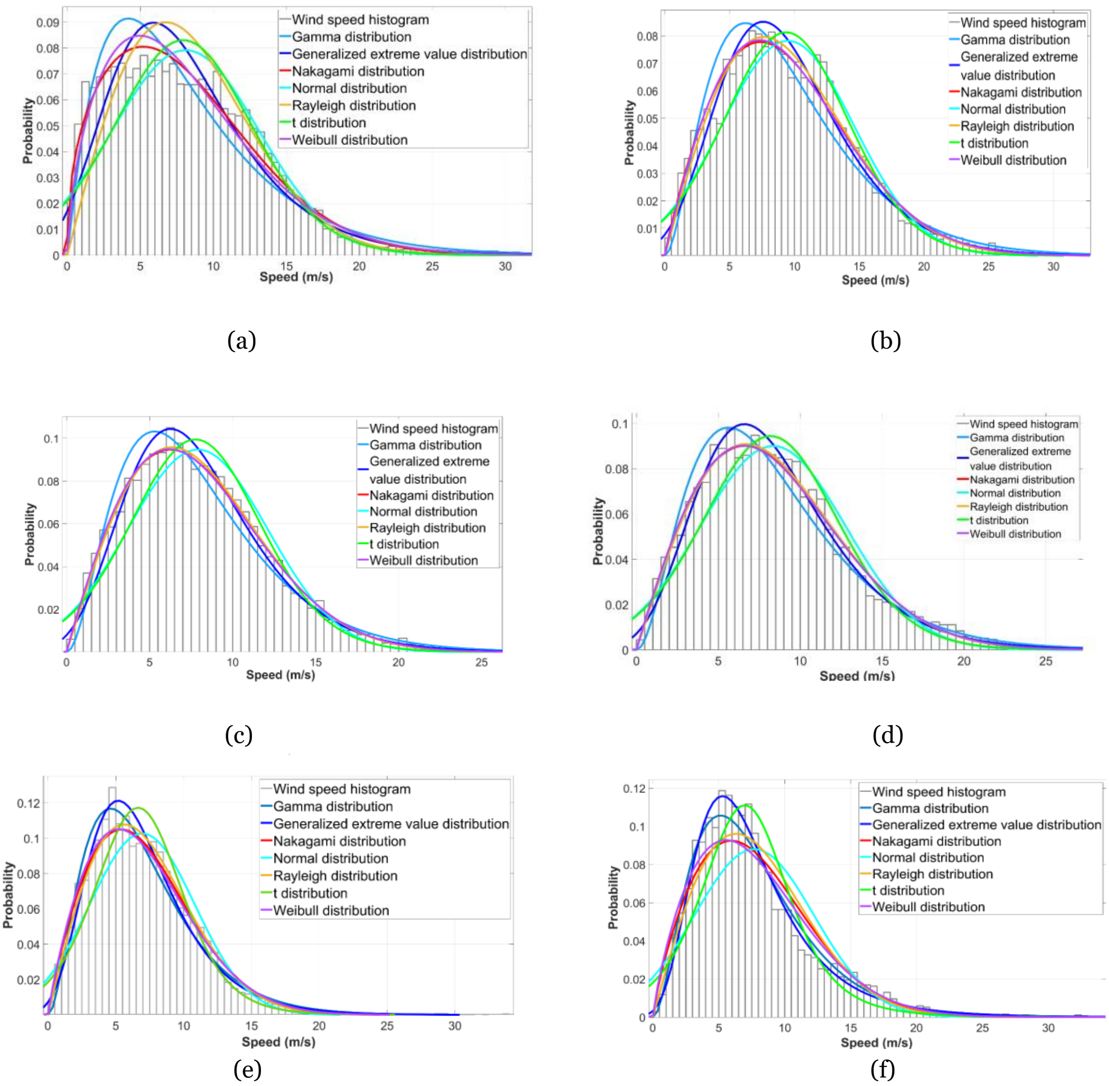


Fig 1. The estimated PDFs curve graphs for NWP model data of five sites and measurements from Fakken wind park (NWP: (a) Nygårdsfjellet, (b) Raggovidda, (c) Kjøllefjord, (d) Havøygavlen, (e) Fakken), Measurements: (f) Fakken)

5.2. K-S test

The K–S test is a rigorous statistical test. Passing this test indicates that there is no statistically significant difference between the PDFs of original data and ideal distributions. The null hypothesis in the present study was that wind speed data fit a mentioned ideal distribution; however, they also could not originate from such an ideal distribution. The significance level was set at 1%, and the results of the K–S test are given in Table 4. 'Pass' means that the K–S test did not reject the null hypothesis, and 'Fail' indicates that the K–S test rejected the null hypothesis.

Table 4. The result of the K-S test

Wind park	**Gamma**	**GEV**	**Nakagami**	**Normal**	**Rayleigh**	***t***	**Weibull**
Nygårdsfjellet	Fail	Fail	Fail	Fail	Fail	Fail	Fail
Raggovidda	Fail	Fail	Pass	Fail	Fail	Fail	Pass
Kjøllefjord	Fail	Fail	Pass	Fail	Pass	Fail	Pass
Havøygavlen	Fail	Pass	Pass	Fail	Pass	Fail	Pass
Fakken (NWP)	Fail	Fail	Fail	Fail	Fail	Fail	Fail
Fakken (MEASURE)	Fail	Pass	Fail	Fail	Fail	Fail	Fail

As is shown, none of the distributions could pass all the K–S tests at the 1% significance level. In addition, the Gamma, normal and t distributions failed the tests in all cases. Meanwhile, the Nakagami and Weibull distributions passed the test for three of the NWP wind data sets, while no distributions passed the tests for Nygårdsfjellet. Regarding the comparison of the PDF modelling between the NWP and observed wind speed of Fakken, all distributions failed the tests for Fakken (NWP), and only the GEV distribution passed the test for Fakken (MEASURE). Therefore, the different probabilistic models each have particular strengths that vary according to wind park and data types.

5.3. Overall wind speed PDF modelling

Table 5 shows the calculated parameters by MLE of different PDF models.

Table 5. The parameters of fitted PDFs

	Gamma	GEV	Nakagami	Normal	Rayleigh	t	Weibull
Nygardsfjellet	2.07; 3.91	5.80; 4.10; -0.02	0.72; 90.93	8.10; 5.04	6.74	17.87	1.62; 9.02
Raggovidda	2.90; 3.27	7.26; 4.10; -0.07	0.94; 116.08	9.50; 5.10	7.62	20.66	1.93; 10.69
Kjøllefjord	3.04; 2.60	6.04; 3.10; -0.05	0.97; 80.15	8.10; 5.04	6.33	16.67	1.96; 8.91
Havoygavlen	3.07; 2.71	6.37; 3.10; -0.05	0.98; 89.14	8.34; 4.43	6.68	16.84	1.96; 9.40

Fakken(NWP)	2.99; 2.32	5.18; 3.10; 0.00	0.94; 63.37	6.95; 3.89	5.63	7.48	1.87; 7.83
Fakken(MEASURE)	3.01; 2.56	5.54; 3.10; 0.08	0.90; 79.62	7.69; 4.53	6.31	4.26	1.80; 8.68

Note: The parameters are shown with the form in Table 2 in order corresponding to each PDF.

The overall MAE of wind speed PDF fitting for the NWP model from five sites and measurements from Fakken is given in Fig. 2. For the NWP wind speed data, the Nakagami distribution generally had a lower MAE than the other distributions. One exception to this is Havøygavlen, in which the Rayleigh distribution performed the best. The normal t distributions had the worst performance in terms of MAE. For the Nakagami distributions for NWP wind speed from different wind parks, the MAEs of Kjøllefjord and Fakken (which are characterised by rougher terrain) were lower compared with the other wind parks. For the observed wind speed data fitting of Fakken, the GEV distribution had the lowest MAE; here, the edge was even more significant than the Nakagami distribution for Fakken NWP data modelling. In addition, the overall MAE of Fakken measured wind speed modelling was much larger than for the NWP data of Fakken.

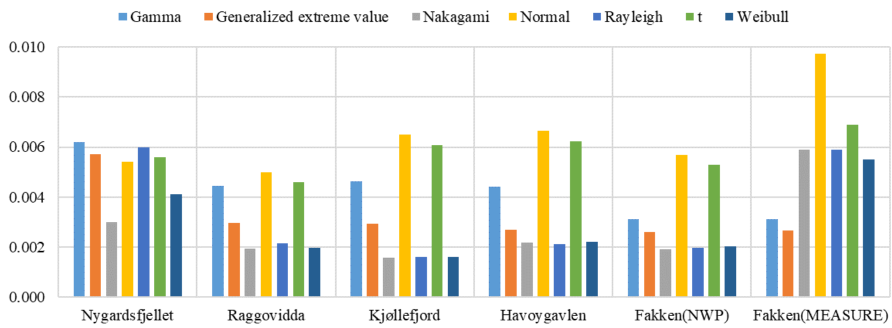


Fig 2. The overall MAE of wind speed PDFs for NWP data from five sites and measurements from Fakken

The overall RMSE of the overall wind speed PDF for the NWP model of five sites and measurements of Fakken wind park is displayed in Fig. 3. In relation to NWP wind speed data, the Nakagami and Rayleigh distributions showed a low RMSE between the histogram and parameterised PDFs, except for Nygårdsfjellet. The overall RMSE of the normal and t distributions was relatively high. Kjøllefjord had the lowest RMSE in the Nakagami and Rayleigh distribution. In terms of the RMSE of wind speed measured data from Fakken, the results were similar to the overall MAE results.

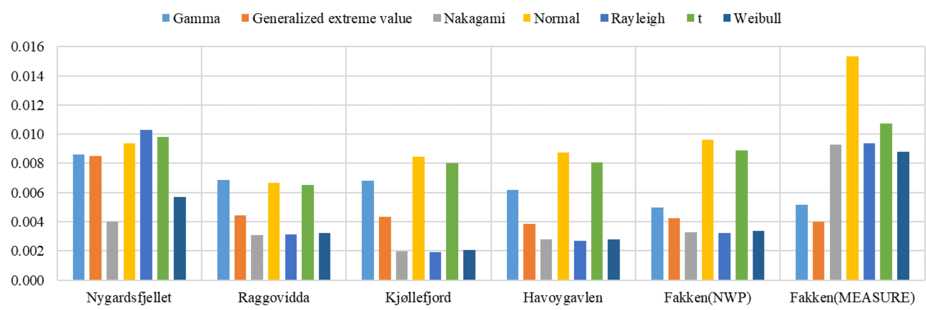


Fig 3. The overall RMSE of wind speed PDFs for NWP data from five sites and measurements from Fakken

Friedman tests for the overall MAE and RMSE of wind speed PDF modelling for the NWP data from five sites were conducted to determine whether there were statistical differences between different probability distribution modelling approaches (effect of distributions) and whether there were statistical differences in the probabilistic modelling results for different wind parks (effect of parks). All the *p*-values surpassed the confidence level of 0.01; therefore, the Friedman test's null hypothesis was not rejected. The results are shown in Table 6.

Table 6. The *p*-values of the Friedman test for overall wind speed modelling

	Effect of distributions	Effect of parks
MAE	0.0011	0.0525
RMSE	0.0014	0.0029

5.4. Interval wind speed PDF modelling

The MAE of interval wind speed PDF is shown in Fig. 4. The results showed some differences from their counterparts in the overall modelling. For the NWP wind speed data, the optimal for Nygårdsfjellet was obtained with the Rayleigh distribution. The Weibull distribution had a slight advantage over the Nakagami and Rayleigh distributions, while the normal distribution showed the worst MAE performance on the whole. The MAE of the Weibull distributions for Kjøllefjord were the smallest out of the five wind parks. Regarding the distributions of measured wind speed of Fakken, the overall MAE was much larger than for the Fakken NWP data; further, the GEV produced the lowest MAE.

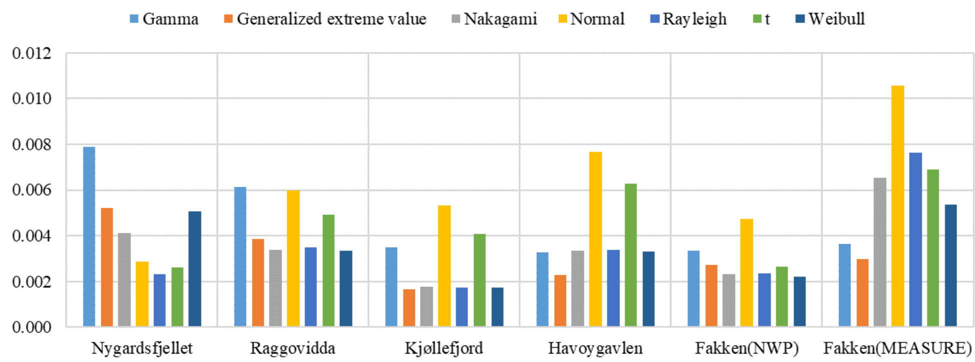


Fig 4. The MAE of interval wind speed PDFs for NWP data from five sites and measurements from Fakken

The RMSE of the interval wind speed PDF is shown in Fig. 5. The results were similar to the MAE evaluation of interval modelling. For the NWP wind speed data, the Rayleigh distribution was superior to other distributions for Nygårdsfjellet. The Nakagami, Rayleigh and Weibull distributions had almost the same RMSEs for the remaining four wind parks, while for the RMSE of the observed data from Fakken, the GEV distribution still won.

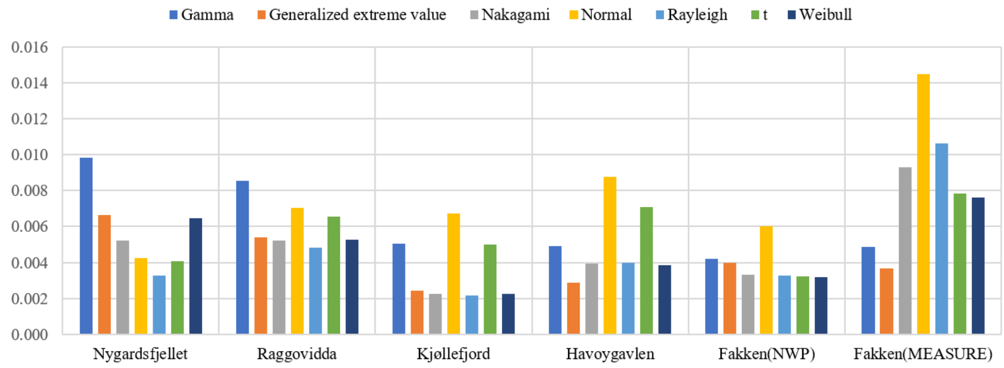


Fig 5. The RMSE of interval wind speed PDFs for NWP data from five sites and measurements from Fakken

Similarly, differences in interval wind speed from NWP probabilistic modelling between wind parks were tested, and the results are given in Table 7. All *p*-values exceeded the confidence level of 0.01, which suggests that there are statistical differences between different probability distribution modelling methods and in the probabilistic modelling of different wind parks.

Table 7. The *p*-values of the Friedman test for interval wind speed modelling

	Effect of distributions	Effect of parks
MAE	0.0717	0.05
RMSE	0.0156	0.0134

5.5. Discussion

In summary, the Nakagami distribution is recommended as the preferred model for the PDF of NWP wind speed data, as it showed excellent and consistent performance. The Nakagami and Weibull distributions could generally capture essential characteristics of the historical distributions of wind speed for both NWP model data by K–S tests. The GEV distribution could describe the statistics of the observed wind data in the examples we used. Moreover, PDF modelling for the NWP wind speed was more accurate compared with actual measurements of wind speed.

In terms of evaluating the NWP wind speed, for the overall wind speed PDF modelling performance, the Nakagami and Weibull distributions showed a good fit for all five wind parks' overall PDFs of NWP wind speed data. In comparison, the Rayleigh distribution provided a favourable overall fit for all except Nygårdsfjellet. The Nakagami and Rayleigh distributions also performed excellently for the wind speed interval modelling. Generally, we made a more precise PDF fitting for NWP wind speed data from Kjøllefjord than for other wind parks both in overall and interval wind speed modelling. This was unexpected, as Kjøllefjord has the highest RIX (10 to 20) of all of them. In addition, Havøygavlen and Fakken, with RIXs (5 to 10), were also fitted better than Nygårdsfjellet and Raggovidda with RIXs (0 to 5). Further research is needed because it is generally thought that the more complex the terrain is, the more difficult it is to use NWP to forecast the wind speed.[38]

For the actual observed wind data from Fakken, the GEV distribution was superior to all other distributions both in overall and interval wind speed modelling and should be used to assess wind speed in this area. The differences between NWP wind data and real measurements of Fakken can be summarised as follows. First, referring to Table 2, the observed speed had a higher mean value, standard deviation, coefficient of variation and skewness, though lower kurtosis meant that the measurements varied more from the normal distribution and had a lighter distribution tail than the NWP data. Second, the best distributions were the Nakagami and Generalised extreme value distribution, respectively. The Weibull distribution, which is typically used for wind speed modelling, was inferior to these two methods in our cases.

6. Conclusions

The statistical characteristics of wind speed are essential for the practical assessment of wind energy potential and the sustainable design of wind parks. In the present study, we concentrated on probabilistic modelling of NWP wind speed for five wind parks in the Norwegian Arctic region and one observed wind speed for one of them. Our results are based on one year of data, and a longer period is needed to conduct a wind resource assessment of a potential wind park site. Using longer time series would provide a better estimate of the wind speed distribution for NWP and measurements and a better understanding of rare extreme high wind events. The results of the present study indicated that, for wind resource assessments in complex terrain, the Nakagami and Generalised extreme value distributions are recommended as the preferred models for the PDF of NWP and observed wind speed, respectively, as they showed excellent and consistent performance. In addition, the probabilistic models that reasonably describe interval wind speed differ from those of overall wind speed due to the nature of the wind: the former corresponds more to the right-side properties of the probability distribution functions.

Based on the results of this study, the following policy recommendations are provided:

1. Different probabilistic modelling approaches should be considered when conducting wind resource potential assessments to achieve more accurate estimations.
2. The wind speeds of neighbouring regional wind parks are characterised by similarities and synergies partly due to the probabilistic models that accurately describe them are identical. But in wind engineering reality, Topography, meteorology, turbine selection and layout etc. all affect the power generation of a wind park. Therefore, the possibility of simultaneous intermittency of these wind parks must be considered when exploiting wind power in the area. Reasonable compensations for other energy sources are required.
3. Compared with observed wind speeds, numerical predicted speeds can be better described by probabilistic models; therefore, when using numerical meteorology to assess wind resources, more consideration should be given to extreme wind events. Some allowance may be made for errors in wind energy project development.

References

1 Zeng, P., Sun, X. & Farnham, D. J. Skillful statistical models to predict seasonal wind speed and solar radiation in a Yangtze River estuary case study. *Scientific reports* **10**, 1-11 (2020).

2 Nazir, M. S., Ali, N., Bilal, M. & Iqbal, H. M. Potential environmental impacts of wind energy development: A global perspective. *Current Opinion in Environmental Science & Health* **13**, 85-90 (2020).

3 Association, E. W. E. *Wind energy-the facts: a guide to the technology, economics and future of wind power*. (Routledge, 2012).
4 Yuan, J. Wind energy in China: Estimating the potential. *Nature Energy* **1**, 1-2 (2016).
5 Wang, J., Hu, J. & Ma, K. Wind speed probability distribution estimation and wind energy assessment. *Renewable and sustainable energy Reviews* **60**, 881-899 (2016).
6 Jain, P. *Wind energy engineering*. (McGraw-Hill Education, 2016).
7 Bilal, M., Birkelund, Y., Homola, M. & Virk, M. S. Wind over complex terrain–Microscale modelling with two types of mesoscale winds at Nygårdsfjell. *Renewable Energy* **99**, 647-653 (2016).
8 Safari, B. & Gasore, J. A statistical investigation of wind characteristics and wind energy potential based on the Weibull and Rayleigh models in Rwanda. *Renewable Energy* **35**, 2874-2880 (2010).
9 Akdağ, S., Bagiorgas, H. & Mihalakakou, G. Use of two-component Weibull mixtures in the analysis of wind speed in the Eastern Mediterranean. *Applied Energy* **87**, 2566-2573 (2010).
10 Ozay, C. & Celiktas, M. S. Statistical analysis of wind speed using two-parameter Weibull distribution in Alaçatı region. *Energy Conversion and Management* **121**, 49-54 (2016).
11 Aries, N., Boudia, S. M. & Ounis, H. Deep assessment of wind speed distribution models: A case study of four sites in Algeria. *Energy Conversion and Management* **155**, 78-90 (2018).
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 Conversion and Management* **119**, 101-108 (2016).
13 Ayodele, T., Jimoh, A., Munda, J. & Agee, J. Wind distribution and capacity factor estimation for wind turbines in the coastal region of South Africa. *Energy Conversion and Management* **64**, 614-625 (2012).
14 Gualtieri, G. & Secci, S. Extrapolating wind speed time series vs. Weibull distribution to assess wind resource to the turbine hub height: A case study on coastal location in Southern Italy. *Renewable Energy* **62**, 164-176 (2014).
15 Allouhi, A. *et al.* Evaluation of wind energy potential in Morocco's coastal regions. *Renewable and Sustainable Energy Reviews* **72**, 311-324 (2017).
16 Jiménez, P., Dudhia, J. & Navarro, J. On the surface wind speed probability density function over complex terrain. *Geophysical research letters* **38** (2011).
17 Birkelund, Y., Alessandrini, S., Byrkjedal, Ø. & Monache, L. D. Wind power predictions in complex terrain using analog ensembles. (2018).
18 Collins, S. N. *et al.* Grids in numerical weather and climate models. *Climate change and regional/local responses* **256** (2013).
19 Bremnes, J. B. & Giebel, G. (The Norwegian Meteorological Institute, 2017).
20 Masseran, N., Razali, A. & Ibrahim, K. An analysis of wind power density derived from several wind speed density functions: The regional assessment on wind power in Malaysia. *Renewable and Sustainable Energy Reviews* **16**, 6476-6487 (2012).
21 Burgin, T. The gamma distribution and inventory control. *Journal of the Operational Research Society* **26**, 507-525 (1975).
22 Hosking, J. R. M., Wallis, J. R. & Wood, E. F. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* **27**, 251-261 (1985).
23 Nakagami, M. in *Statistical methods in radio wave propagation* 3-36 (Elsevier, 1960).
24 Rubio, L., Reig, J. & Cardona, N. Evaluation of Nakagami fading behaviour based on measurements in urban scenarios. *AEU-International Journal of Electronics and Communications* **61**, 135-138 (2007).
25 Lyon, A. Why are normal distributions normal? *The British Journal for the Philosophy of Science* **65**, 621-649 (2014).
26 Balakrishnan, N. Approximate MLE of the scale parameter of the Rayleigh distribution with censoring. *IEEE Transactions on Reliability* **38**, 355-357 (1989).
27 Papoulis, A. & Pillai, S. U. *Probability, random variables, and stochastic processes*. (Tata McGraw-Hill Education, 2002).
28 Rigby, R. A. & Stasinopoulos, D. M. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **54**, 507-554 (2005).
29 Rinne, H. *The Weibull distribution: a handbook*. (CRC press, 2008).

30 Saleh, H., Aly, A. A. E.-A. & Abdel-Hady, S. Assessment of different methods used to estimate Weibull distribution parameters for wind speed in Zafarana wind farm, Suez Gulf, Egypt. *Energy* **44**, 710-719 (2012).

31 Chang, T. P. Performance comparison of six numerical methods in estimating Weibull parameters for wind energy application. *Applied Energy* **88**, 272-282 (2011).

32 Myung, I. J. Tutorial on maximum likelihood estimation. *Journal of mathematical Psychology* **47**, 90-100 (2003).

33 Gibbons, J. D. & Chakraborti, S. *Nonparametric Statistical Inference: Revised and Expanded*. (CRC press, 2014).

34 Justel, A., Peña, D. & Zamar, R. A multivariate Kolmogorov-Smirnov test of goodness of fit. *Statistics & Probability Letters* **35**, 251-259 (1997).

35 Hollander, M., Wolfe, D. A. & Chicken, E. *Nonparametric statistical methods*. Vol. 751 (John Wiley & Sons, 2013).

36 Marsaglia, G., Tsang, W. W. & Wang, J. Evaluating Kolmogorov's distribution. *Journal of statistical software* **8**, 1-4 (2003).

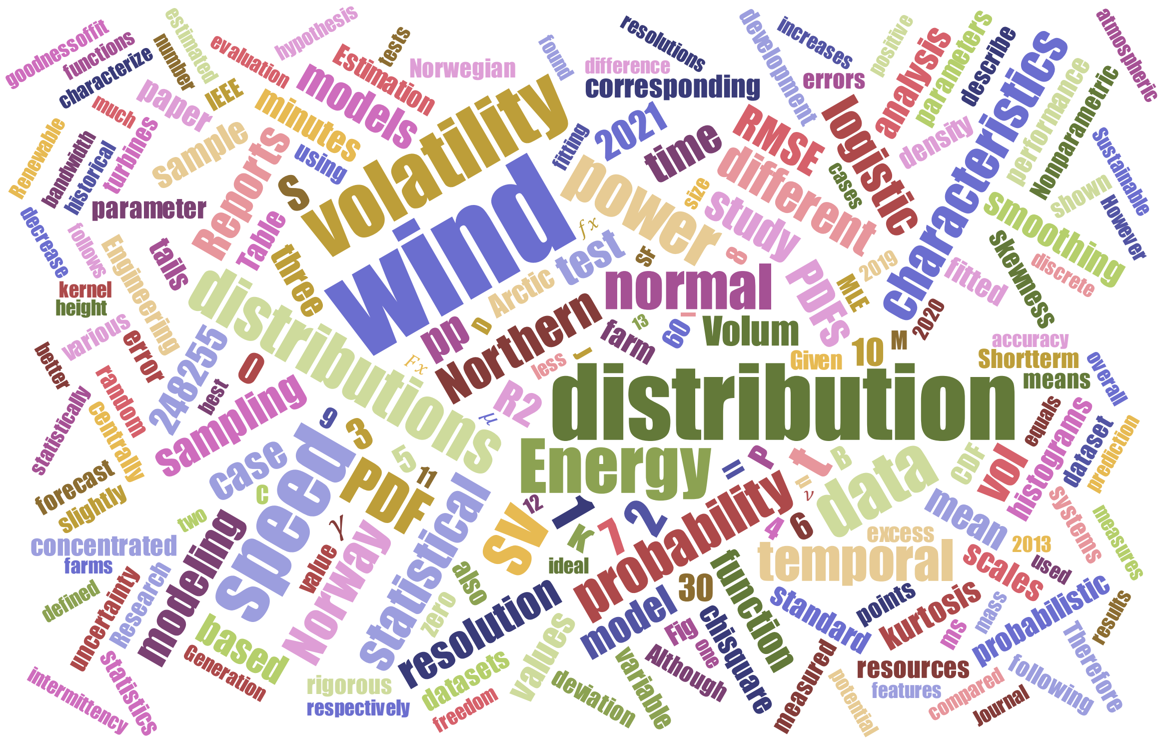
37 Li, L. *et al.* A half-Gaussian fitting method for estimating fractional vegetation cover of corn crops using unmanned aerial vehicle images. *Agricultural and Forest Meteorology* **262**, 379-390 (2018).

38 Cassola, F. & Burlando, M. Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output. *Applied energy* **99**, 154-166 (2012).

Paper I. B.

Probability distributions for wind speed volatility characteristics: A case study of Northern Norway

https://doi.org/10.1016/j.egyr.2021.07.125



Tmrees, EURACA, 28 to 30 May 2021, Athens, Greece

Probability Distributions for Wind Speed Volatility Characteristics: A Case Study of Northern Norway (Postprint)

Hao Chen[a, *], Stian Normann Anfinsen[b], Yngve Birkelund[b], Fuqing Yuan[a]

[a]*Department of Technology and Safety,*
[b]*Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø 9019, Norway*

Abstract

The Norwegian Arctic is rich in wind resources. The development of wind power in this region can boost green energy and also promote local economies. In wind power engineering, it is a tremendous advantage to base projects on a sound understanding of the intrinsic properties of wind resources in an area. Wind speed volatility, a phenomenon that strongly affects wind power generation, has not received sufficient research attention. In this paper, a framework for studying short-term wind speed volatility with statistical analysis and probabilistic modeling is constructed for an existing wind farm in Northern Norway. It is found that unlike the characteristics of wind power volatility, wind speed volatility cannot be described by the normal distribution. The reason is that even though the probability distribution of wind speed volatility is centrally symmetric, it is much more centrally concentrated and has thicker tails. After comparing three distributions corresponding to different sampling periods, this paper suggests utilizing the *t* distribution, with average modeling RMSE less than 0.006 and R^2 exceeding 0.995 and with the best modeling scenario of temporal resolution, the 30 mins has an RMSE of 0.0051 and an R^2 of 0.997, to more accurately and effectively explore the fluctuating characteristics of wind speed.

Keywords: Wind energy; Wind speed volatility; Statistical analysis; Probability distribution; Arctic

Abbreviations

PDF	probability density function
CDF	cumulative distribution function
SV	wind speed
SP	wind speed volatility
γ	skewness
κ	excess kurtosis
MLE	maximum likelihood estimation
RMSE	root mean square error
R^2	coefficient of determination

1. Introduction

As an alternative to fossil fuels, wind energy has received increasing attention worldwide because of its abundant availability, widespread dispersal, and potential financial support [1]. Norway owns some of the best wind energy resources in Europe [2]. It has enormous potential for wind power generation, especially in its northern and Arctic regions.

Assessing potential wind resources—typically evaluated by measured and modeled wind speed and direction through a year or more at a certain location—are critical for evaluating the feasibility and sustainability of a wind energy project [3]. Wind is a phenomenon involving air movement and relates to the atmospheric motion state. Changes in wind characteristics are closely related to the circulation of energy and matter in the atmosphere. The most noticeable difference between wind energy and conventional energy is the volatility, stochasticity, intermittency, and uncontrollability of the former [4]. The changes in wind speed are affected by long-term atmospheric motion and micro-scale atmospheric turbulence caused by many surface factors. These cause the wind to show strong instantaneous volatility in time and space. Due to the uncertainty and intermittency of wind, wake effects between wind turbines, and the cubic relationship between wind speed and the wind turbine-generated power, a small change of wind speed can be significantly amplified in the output wind power. The random volatility of wind is regarded as an adverse factor for wind energy [5]. This intermittency brings severe challenges to the power system's safety, power quality, and the balance of power supply and demand. Therefore, studying the volatility characteristics of wind is of great significance for improving wind power forecasting accuracy, scenario generations, and overcoming the adverse effects of wind power integration in the grid [6].

However, the typical wind energy assessment methodology lacks tools to characterize wind speed volatility on sites. The volatility analysis offers additional information about wind. The wind has different volatility characteristics at different temporal scales. Although the wind has certain seasonal and diurnal characteristics, there is no fixed volatility amplitude and cycle; its volatility has no clear rules to follow.

The probability density function (PDF) is an effective quantification to describe wind randomness and uncertainty [7]. Much research has used the probability density function in wind engineering [8]. However, most of the research concerns evaluation of historical wind speed distribution. To illustrate, Mahmood and colleagues [9] used the Weibull distribution to assess wind speed data from a site in Iraq successfully.

Studies centered on statistical analyses of volatility in wind energy, and those who exist have mainly considered wind power volatility directly are few and far between, although a handful exists. For instance, Lange [10] analyzed the uncertainty in wind power prediction using the statistical distributions and found that wind speed prediction error is normally distributed. Bludszuweit [11] looked into the statistical distributions of wind power errors forecasted by the persistence model. It proposed an indirect algorithm based on the Beta distribution based on one-year measured data from two different wind farms. Zhang [12] presented a versatile distribution for fitting wind power predictive errors and compared the distribution with benchmarking distributions of normal and Beta. Inspired by probability distributions of wind power volatility, it is also possible to use statistics to analyze the wind speed volatility using classical ideal distribution functions to model the histogram of the wind volatility and capture its nature.

This paper uses different probability density functions and skewness and kurtosis moments to characterize short-term wind speed volatility at various temporal scales for a wind farm in Northern Norway. The statistical modeling of volatility assists in documenting wind's internally volatile features, especially for the wind in a cold climate and complex terrain.

2. Data preparation

This paper draws on data from a wind power station located in the Norwegian Arctic, whose coordinates are 70°5'56" N, 20°3'54" E, and its designed capacity is 54 MW. The hub height of the turbines is 80 m above the ground, and the rotor diameter is 90 m. The farm has eighteen Vestas V90-3.0 3.0 MW turbines with 45m long rotor and the hub height is 80m above the ground. The wind farm is surrounded by hills and fronts a fjord. The wind park company provides measurements, taken by the wind mast with the same height of turbines, of wind speed. The original wind speed data are from 0:00 on 1st January 2017 to 23:00 on 31st December 2017 with 10 minutes temporal resolution. The number of measured data points is 52,560. Wind speed data with a reduced temporal resolution of 30 minutes and 60 minutes are obtained by interpolations. The size of the dataset with 30 minutes and 60 minutes resolution is 17,520 and 8,760 data points, respectively. The wind *Speed Volatility (SV)* is calculated as the first-order differential by Eqs. (1):

$$SV_i = SP_i - SP_{i-1} \tag{1}$$

where SP_i and SP_{i-1} are wind speed at time t and one temporal resolution before t_i.

3. Methodology

The sample skewness (γ) and sample excess kurtosis (κ) are common shape-parameters that describe the historical distributions of variables, and they are defined as:

$$\gamma = T^{-1} \sum_{t=1}^{T} (X_t - \bar{X})^3 / s^3 \tag{2}$$

$$\kappa = T^{-1} \sum_{t=1}^{T} (X_t - \bar{X})^4 / s^4 - 3 \tag{3}$$

where T is the size of the data sample, $\bar{X}$ is the sample mean, and s is the sample standard division. γ measures whether the PDF of a random variable "leans" to one side of the mean. A distribution is left-skewed when γ is negative and right-skewed when γ is positive. κ measures the "peakedness" of a distribution. The so-called excess kurtosis defined in Eq. (3) is measured relative to the normal distribution, which attains a value of κ=0. Therefore, excess kurtosis is a measure of departure from normality and reflects the sharpness of the peak [13]. A distribution is leptokurtic when $\kappa > 0$, indicating that the PDF is sharper and steeper than the normal distribution, and it is platykurtic when $\kappa < 0$.

The PDF of a random variable is a statistical model that describes the probability of occurrence of this variable at a specific point in each observation interval. The cumulative distribution function (CDF) specifies the probability that the variable is less than or equal to a specific value [14]. In this section, three commonly used ideal PDFs are chosen as the candidates for modeling the SF.

For the normal distribution, its PDF (4) and CDF (5) are expressed by:

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{4}$$

$$F(x; \mu, \sigma) = \frac{1}{2}\left[1 + erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right] \tag{5}$$

where μ is the mean, σ is the standard division, and *erf* (.) is the error function.

The logistic distribution resembles the normal distribution in shape but has heavier tails (higher κ). The PDF (6) and CDF (7) of the logistic distribution are given [15], respectively, by:

$$f(x;\mu,s) = \frac{e^{-(x-\mu)/s}}{s\left(1+e^{-(x-\mu)/s}\right)^2} \tag{6}$$

$$F(x;\mu,s) = \frac{1}{1+e^{-(x-\mu)/s}} \tag{7}$$

where μ is a location parameter and s is a scale parameter. The mean equals μ, and the variance is $s^2 \pi^2/3$.

The PDF (8) and CDF (9) of the t distribution are determined via the following functions [16]:

$$f(x;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\nu\pi}} \frac{1}{\left(1+\frac{x^2}{\nu}\right)^{\frac{\nu+1}{2}}} \tag{8}$$

$$F(x;\nu) = \int_{-\infty}^{x} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\nu\pi}} \frac{1}{\left(1+\frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}} dt \tag{9}$$

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

Since histograms are discrete distributions, a nonparametric method of simulating distributions based on the data itself, the kernel distribution, can approximate discrete historical distributions to the empirical distribution of samples taken at infinitely small intervals. Figuratively, it is called *smoothing PDFs* and is determined by a smoothing function and a bandwidth. In this study, the smoothing function is the Gaussian function, and the bandwidth values 0.025, which can extract wind speed information with high precision and without adding sampling noise.

3.1. Parameter estimation

The PDF parameter estimation means an ideal probability distribution model can statistically describe the distribution of SV data. The parameters of the model are estimated by training the SV data with proper estimation approaches. This study uses the Maximum Likelihood Estimation (MLE) approach to determine parameters for the above three PDFs.

4. Experiments

The procedure for modeling the PDF of SV at different temporal scales is illustrated in Fig. 1. The raw wind speed data are interpolated and calculated by Eq. (1) to create SF data sequences for different temporal scales. These data are then tested for their normality, and their histograms are plotted. Moreover, their *smoothing PDFs* are created by the kernel distribution. Then, fitted distribution models corresponding to all SF datasets on different temporal scales are created, whose parameters are obtained with the MLE method. Finally, the fitted PDF models are tested with the goodness-of-fit and compared with the corresponding smoothing PDFs.

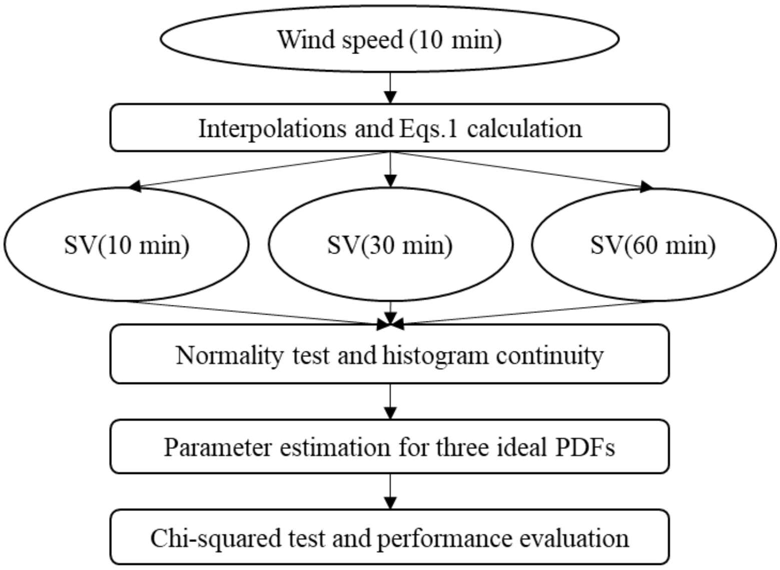


Fig. 1. Procedure for the SV probabilistic modeling.

Pearson's chi-square goodness-of-fit test for PDF models is a nonparametric test that evaluates how likely a data sample has been drawn from a given PDF [17]. The chi-square test divides data into k bins and defines the following null hypothesis: H_o: {X1, X2, ⋯, X_n } follows the given probability distribution. The alternative hypothesis is: H_1: {X1, X2, ⋯, X_n } do not follow this distribution. The test statistic is defined by Eqs. (10):

$$\chi^2 = \sum_{i=1}^{k} (O_i - E_i)^2/E_i \tag{10}$$

where O_i is the observed count and E_i is the expected count for bin i based on the hypothesized PDF.

To evaluate the performance of different PDFs for SV modeling, the Root Mean Square Error (RMSE) and the coefficient of determination (R^2) are applied to calculate the probability density difference between smoothing PDFs and corresponding fitted PDF models. RMSE is a negatively oriented metric, meaning that smaller values indicate better fitting performance. Meanwhile, the second is positively oriented, and its range is zero between one.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(smoothing_i - modeling_i)^2}{n}} \tag{11}$$

$$R^2 = 1 - \frac{SS_{\mathrm{res}}}{SS_{\mathrm{tot}}} \tag{12}$$

Where n is the total number of sampling by the kernel function with 0.025m/s bandwidth that is related to SV ranging from -10m/s to 10m/s, and it equals 800. SS_{res} is the sum of squares of residuals (deviations fitted from smoothing PDFs based on histograms) and SS_{tot} is the total sum of squares (overall squared differences between the smoothing PDF values at the sampling points and their averages).

5. Results and discussion

We use three different PDFs, the normal distribution, the logistic distribution, and the t distribution, to model the volatility of wind speed over various temporal intervals for the wind farm in Northern Norway. The results are presented as follows.

5.1. Statistics for SV data

The descriptive statistics for SV data are shown in Table 1.

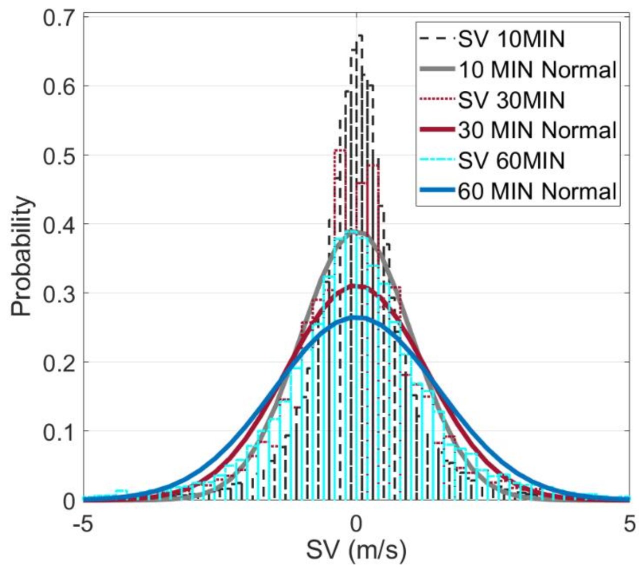
Table 1. The statistics of SV data at different temporal scales.

Temporal resolution (minutes)	Mean (m/s)	Standard Deviation (m/s)	Min (m/s)	Max (m/s)	Skewness	Kurtosis
10	0.0000	1.0200	-9.6000	16.2000	0.3741	8.2444
30	-0.0003	1.2808	-9.7000	11.6333	0.2429	5.1468
60	-0.0007	1.5034	-10.4167	12.2833	0.1930	4.4025

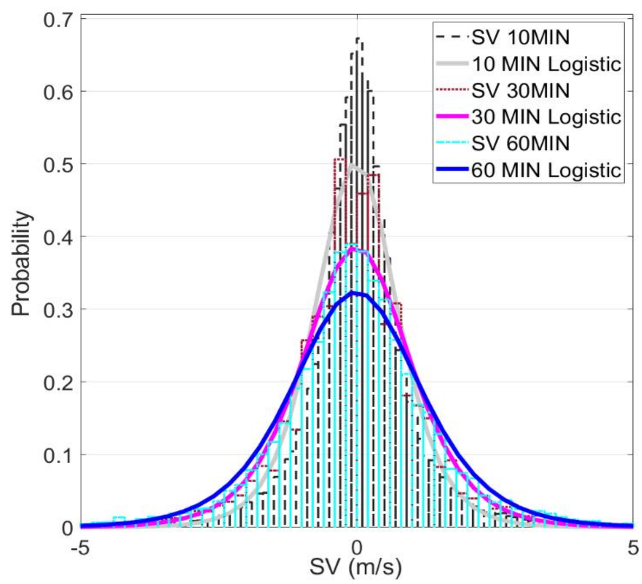
The mean value of the SV data is very close to zero at all temporal resolutions, which indicates that the wind speed volatility is, in general, trendless and oscillates back and forth around the zero points. As the sampling time grows, the SV data standard deviation increases, and their γ and κ decrease. The increase in standard deviation is understandable since SV is more variable over more extended periods. The γ of all three SV datasets is slightly positive, which means the right tails of the distributions are longer than the left ones, and their mass is concentrated slightly to the left. Both γ and κ decrease with time spacing, and so the data become increasingly normal. The negative correlation of the γ with sampling time indicates that the histogram of the SV data becomes more symmetrical as the time spacing increases. The three SV datasets have positive κ, which shows that all of them are leptokurtic and morphologically steeper or thicker tails than the normal distribution. Large κ values can occur in two situations: the probability mass is concentrated near the mean, and occasionally, there are some data in the dataset that are away from the mean, or the mass of probability is concentrated at the tails of the distribution. κ values that increase with temporal resolution also illustrate the decline in the concentration and the size of extreme values away from the means of SV datasets. Therefore, based on the above analysis, it is reasonable to assume that merely using the normal distribution to describe SV is inaccurate.

5.2. PDF modeling fitting and test

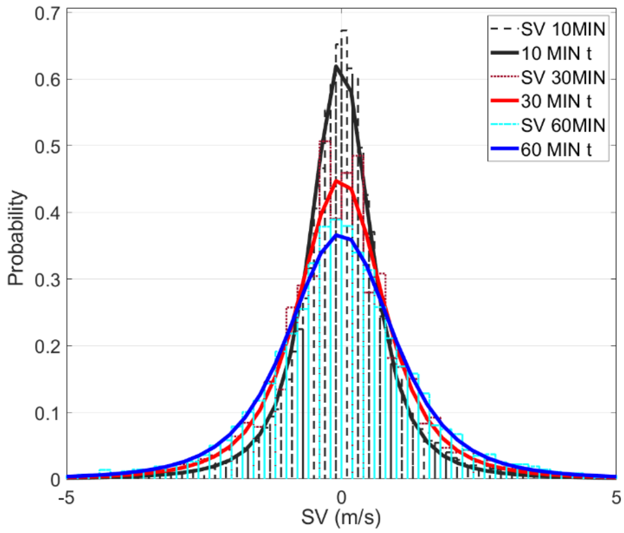
Fig. 2 shows histograms of SV data and fitted model PDFs with parameters that have been estimated by the MLE approach.


SV 10MIN
10 MIN Normal
SV 30MIN
30 MIN Normal
SV 60MIN
60 MIN Normal
Probability
SV (m/s)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-5
0
5

(a)


SV 10MIN
10 MIN Logistic
SV 30MIN
30 MIN Logistic
SV 60MIN
60 MIN Logistic
Probability
SV (m/s)
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-5
0
5

(b)



(c)

Fig. 2. The histograms and estimated PDFs curve graph for SV data ((a) is normal, (b) is logistic, and (c) is t distribution)

From Fig. 2, it can be seen visually that the mode PDF value falls significantly as the sampling time increases. The t and logistic distributions fit the shape of histograms better than the normal distribution for all three temporal resolution cases.

Table 2. The parameters for PDFs and *p* values of the chi-square test (Values less than 10-8 are approximately expressed as zero).

Temporal resolution (minutes)	Normal (μ,σ)	Logistic (μ,s)	t (v)	Normal p	Logistic p	t p
10	(0,1.02)	(0,0.50)	2.50	0	0	0
30	(0,1.28)	(0,0.65)	3.00	0	0	0
60	(0,1.50)	(0,0.77)	3.31	0	0	0.062

The standard deviation of the normal distribution, the scale parameter of logistic distribution, and the degrees of freedom of t distribution are shown in Table 2 to correlate positively with the time resolution, proving that curves of all three distributions become " lower and broader.

Pearson’s chi-square test is a rigorous statistical test. According to the test, it can be concluded whether there exists a statistically significant difference between a theoretical distribution model and the observed frequency distribution of specified discrete events in the data sample. The hypothesis tests areas above section 4.2 and with a significance level of 5%. The p values of chi-square tests are also shown in Table 2. Only the p-value for the t distribution corresponding to the SV data with 30 minutes is above 0.05, indicating that the dataset statistically follows the t distribution with a degree of freedom equals 3.31. Given that rigorous statistical tests do not give a complete picture of the accuracy of probabilistic models. We will introduce quantitative analysis to evaluate these models in the following sub-section.

5.3. Performance evaluation

Real-world data will often have problems with passing a rigorous statistical test. In engineering practice, evaluation metrics from regression analysis are commonly adopted to assess the quality of PDF modeling. The RMSE and R^2 between empirical or smoothing PDFs of SV data and different fitted PDF models for various temporal resolutions are shown in Tables 3.

Table 3. RMSE and R^2 of different PDF models

	RMSE			R^2		
Temporal resolution (min)	normal	logistic	t	normal	logistic	t
10	0.0401	0.0226	0.0058	0.8738	0.9596	0.9971
30	0.0286	0.0150	0.0051	0.9106	0.9748	0.9970
60	0.0236	0.0121	0.0055	0.9241	0.9795	0.9954

It is found that although most of the PDF models do not pass the chi-square test, the R^2 of all logistic and t distribution models surpasses 0.95, which generally means that these PDFs provide a sound fit. Except for the normal distribution, the other two distributions can display probabilistic characteristics of the SV dataset. Regarding performance differences between different PDF models, the t distribution is superior to other distributions for all sampling time datasets in both RMSE and R^2. Almost all PDF curves are highly centrally concentrated and have heavy and long tails, which potentially embodies the risk of wind ramp events. The t distribution satisfactorily embodies these features. Besides, the logistic distribution performs better than the normal distribution in all cases, suggesting that it can also deliver relatively satisfactory probabilistic modeling for describing SV.

Concerning the comparison of various time resolutions, the RMSE and R^2 of normal and logistic distributions respectively decrease and increase with the sampling time. This demonstrates that both PDFs more easily characterize the SV data's statistical distributions with the rising sampling time. Meanwhile, the RMSE and R^2 of the t distribution are very stable and do not fluctuate much with sampling interval slightly volatile features. Overall, the t distribution is proven to be a more desirable probabilistic model to represent wind speed volatility in comparison with the normal logistic distributions.

6. Conclusion

Statistical characterization of wind volatility is vital to effectively conduct practical assessments of wind resources for wind power development. In the present paper, we focus on statistical modeling of wind speed volatility for a wind farm inside the Norwegian Arctic region. Based on the statistical analysis and PDFs modeling results, the following conclusions can be drawn.

The probability distribution of wind volatility is overall centrally symmetrical but quite different from the normal distribution. In our cases, wind volatility is slightly left-skewed and has sharper peaks compared to the normal distribution. However, as the temporal resolution of sampling decreases, its probability distribution becomes closer to the normal distribution. Although most PDF models fail a rigorous nonparametric goodness-of-fit test based on the raw data of complex wind phenomena, the logistic and t distributions deliver R^2 exceeding 0.95 and RMSE approaching zero, suggesting that both distributions provide good characterizations of wind speed short-term volatility in wind energy engineering practice. Moreover, the t distribution has a notable advantage, and its performance is very stable with sampling time. Therefore, this paper recommends explicitly applying the t distribution to modeling wind speed volatility based on our results.

References

[1] S. Samadianfard *et al.*, "Wind speed prediction using a hybrid model of the multi-layer perceptron and whale optimization algorithm," *Energy Reports,* vol. 6, pp. 1147-1159, 2020.

[2] B. Blindheim, "Implementation of wind power in the Norwegian market; the reason why some of the best wind resources in Europe were not utilised by 2010," *Energy policy,* vol. 58, pp. 337-346, 2013.

[3] M. Ferreira, A. Santos, and P. Lucio, "Short-term forecast of wind speed through mathematical models," *Energy Reports,* vol. 5, pp. 1172-1184, 2019.

[4] G. Notton *et al.*, "Intermittent and stochastic character of renewable energy sources: Consequences, cost of intermittence and benefit of forecasting," *Renewable and sustainable energy reviews,* vol. 87, pp. 96-105, 2018.

[5] T. Yang, "Optimal sizing of the hybrid energy storage system aiming at improving the penetration of wind power," in *2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, 2016: IEEE, pp. 2358-2362.

[6] S. Z. Moghaddam, "Generation and transmission expansion planning with high penetration of wind farms considering spatial distribution of wind speed," *International Journal of Electrical Power & Energy Systems,* vol. 106, pp. 232-241, 2019.

[7] J.-N. Li, Y. Qiao, Z.-x. Lu, J. Li, and F. Xu, "Research on statistical modeling of large-scale wind farms output volatilitys in different spacial and temporal scales," *Power System Protection and Control,* vol. 40, no. 19, pp. 7-13, 2012.

[8] C. Jung and D. Schindler, "Wind speed distribution selection–A review of recent development and progress," *Renewable and Sustainable Energy Reviews,* vol. 114, p. 109290, 2019.

[9] F. H. Mahmood, A. K. Resen, and A. B. Khamees, "Wind characteristic analysis based on Weibull distribution of Al-Salman site, Iraq," *Energy Reports,* vol. 6, pp. 79-87, 2020.

[10] M. Lange, "On the uncertainty of wind power predictions—Analysis of the forecast accuracy and statistical distribution of errors," *J. Sol. Energy Eng.,* vol. 127, no. 2, pp. 177-184, 2005.

[11] H. Bludszuweit, J. A. Domínguez-Navarro, and A. Llombart, "Statistical analysis of wind power forecast error," *IEEE Transactions on Power Systems,* vol. 23, no. 3, pp. 983-991, 2008.

[12] Z.-S. Zhang, Y.-Z. Sun, D. W. Gao, J. Lin, and L. Cheng, "A versatile probability distribution model for wind power forecast errors and its application in economic dispatch," *IEEE transactions on power systems,* vol. 28, no. 3, pp. 3114-3125, 2013.

[13] D. Joanes and C. Gill, "Comparing measures of sample skewness and kurtosis," *Journal of the Royal Statistical Society: Series D (The Statistician),* vol. 47, no. 1, pp. 183-189, 1998.

[14] W. Feller, *An introduction to probability theory and its applications, vol 2*. John Wiley & Sons, 2008.

[15] A. Di Crescenzo and B. Martinucci, "A damped telegraph random process with logistic stationary distribution," *Journal of applied probability,* vol. 47, no. 1, pp. 84-96, 2010.

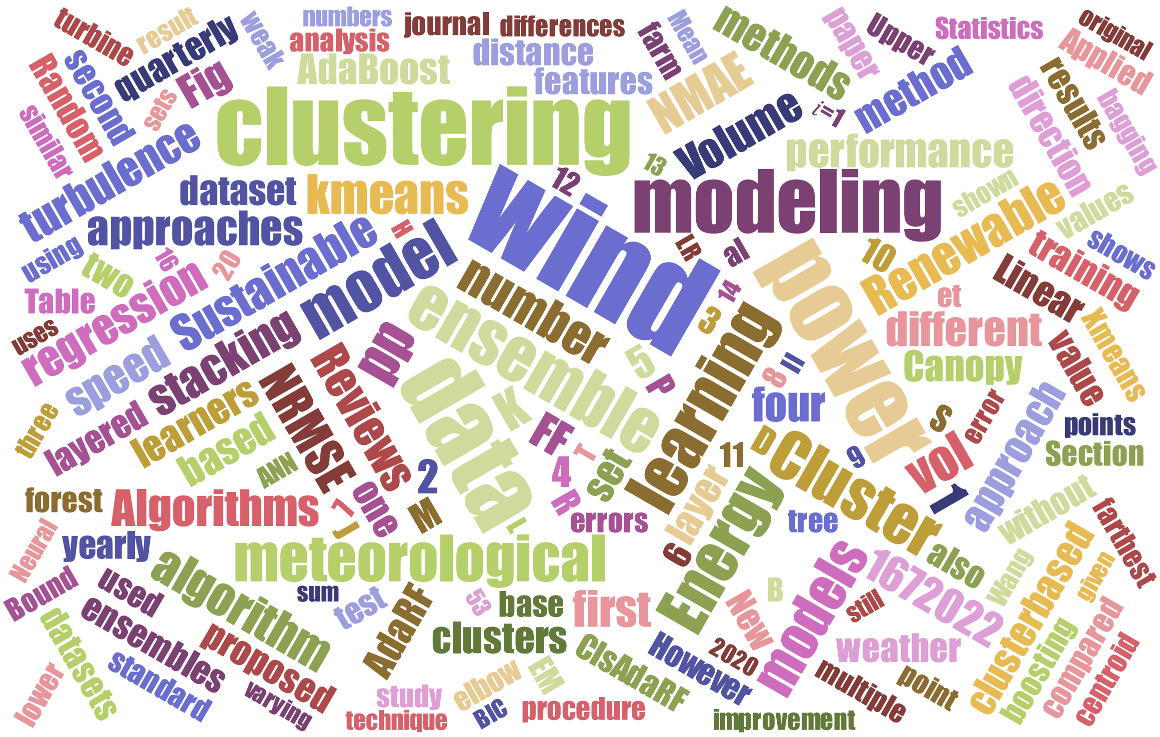
[16] P. Theodossiou, "Financial data and the skewed generalized t distribution," *Management Science,* vol. 44, no. 12-part-1, pp. 1650-1661, 1998.

[17] E. Brodsky and B. S. Darkhovsky, *Nonparametric statistical diagnosis: problems and methods*. Springer Science & Business Media, 2013.

Paper II.

Cluster-based ensemble learning for wind power modeling from meteorological wind data

https://doi.org/10.1016/j.rser.2022.112652



Cluster-based ensemble learning for wind power modeling from meteorological wind data (Postprint)

Hao Chen*

Department of Technology and Safety; Arctic Centre for Sustainable Energy,

UiT The Arctic University of Norway, Tromsø, Norway

* = corresponding author, hao.chen@uit.no

Abstract

Reliable and efficient power modeling from meteorological wind data is vital for optimal implementation and monitoring of wind energy, and it is important for understanding turbine control, farm operational optimization, and grid load balance. Based on the idea that similar wind conditions lead to similar wind powers; this paper constructs a modeling scheme that orderly integrates three types of ensemble learning algorithms—bagging, boosting, and stacking—and clustering approaches to achieve wind power modeling from multiple wind-based meteorological factors in a wind farm. The paper also investigates the applications of different clustering algorithms and methodologies to determine cluster numbers in the modeling. The results reveal that all ensemble models with clustering exploit the intrinsic information in wind data and thus outperform models without clustering by approximately 15% on average in modeling wind power. The model with the best-performing Farthest First clustering is computationally rapid and with an improvement of around 30% compared with the baselines. Given the diversity introduced by clustering algorithms, the power modeling performance is further boosted by about 5% by introducing stacking that fuses ensembles with varying clusters. The proposed modeling framework thus demonstrates promise by delivering efficient and robust performance on the targeted problem.

Highlights

- Systematic demonstration of wind power modeling from weather data
- Construction and comparison of various clustering methods for classifying wind data
- Combination of bagging, boosting, and stacking for modeling
- Multiple wind meteorological characteristics are considered in the modeling

Keywords: wind power modeling, clustering, layered ensemble learning, farthest first algorithm, stacking, Arctic

Abbreviations

NWP	Numerical Weather Prediction
K-means	K-means clustering
EM	Expectation-Maximization clustering
FF	Canopy clustering
Canopy	X-means clustering
X-means	Farthest First clustering
SSE	Sum of Squared Errors
BIC	Bayesian Information Criterion
NMAE	Normalized Mean Absolute Error
NRMSE	Normalized Root Mean Square Error
Bagging	Bootstrap aggregating
Adaboost	Adaptive boosting
REPTREE	Reduced-Error Pruning TREE
AdaRF	Adaboost with Random Forest
LR	Linear Regression
ANN	Artificial three-layer Neural Networks
AdaDT	Adaboost Decision Tree
CoV	Coefficient of Variation
'Cls-' AdaRF, LR, ANN, AdaDT	Two-layer stacking with four clustering methods as the first and AdaRF, LR, ANN, AdaDT as the second, respectively
NCl-AdaRF	Emphasis on AdaRF without a clustering layer (same with AdaRF)
FF-AdaRF	Two-layer stacking with FF clustering and AdaRF

1. Introduction

Wind energy is one of the most commercially viable renewable energy sources, due to its natural abundance and non-reliance on fossil fuels, and has thus become integral to combating climate change. The Global Wind Energy Council estimates that 355 GW of new capacity will be installed between 2020 and 2024, with almost 71 GW of new capacity per year [1]. The rising prevalence of wind energy brings new challenges for planning electricity generation and dispatching grids because wind power varies intermittently and unpredictably with prevailing wind conditions.

Establishing an accurate power model for a wind farm based on the empirical mapping of weather data is important to understand the relationship between wind and wind power generation, which in turn is significant for the safe and stable operation of a wind farm and its economic operation [2]. It is also important for farms to have a non-parametric power curve model that can be applied as a reference profile for the online monitoring generation process [3]. This article aims to conceptualize wind power modeling by using weather data, rather than only wind speed, to compute its corresponding wind power.

1.1. Related works

Driven by progress in computing affordability and capability and algorithmic advances, wind power can increasingly be modeled by physical, statistical, and hybrid methodologies. However, there is still room to improve these models [4].

A few studies have considered meteorological factors in wind power modeling. R. Liu et al. [5] inputted wind speed, wind direction, and air pressure to a power model based on multivariable phase space reconstruction—the similarity of time-series and linear regression—and demonstrated its superiority for forecasting under conditions where wind power series fluctuate considerably. J. Ma et al. [6] used hourly wind speed and direction at the heights of 10 m and 100 m to establish a good performance model through multivariate empirical dynamic modeling. However, this type of research focused on mapping the relationship between weather data and wind power, without examining meteorological data themselves and their potential to improve these models.

Ensemble learning also remains a popular approach to improving modeling, since it can reduce the variance and bias of learners. D. Niu et al. [7] established a wind-speed power model with wavelet decomposition and weighted random forest optimized by the niche immune lion algorithm. The model was subsequently tested in two empirical analyses. Y. Dong et al. [8] processed input data with wavelet packet decomposition and applied a stacking ensemble by evaluating the correlation coefficient between base learners for wind power forecast modeling. The model clearly showed the ensemble edge when the base learners have high accuracy and low correlations with one another. However, these studies were primarily algorithm-oriented and rarely considered wind data's inherent characteristics in detail; meanwhile, their uses of decomposition to handle data increased the modeling time complexity.

Wind has some internal trends that can be understood through data mining approaches. There have been some studies on clustering technique applications in wind power modeling. V. Kushwah et al. [9] found that clusters of time series data showed identical trend components in wind speed data using a cluster-based statistical modeling technique, which showed better performance than other statistical ones. However, purely statistical models can suffer from underfitting problems when dealing with complex data. L. Dong et al. [10] utilized cluster analyses of the Numerical Weather Prediction (NWP) since wind power and corresponding meteorological data have the characteristic of daily similarity. This suggests that the clustering model is useful in the day-ahead modeling of wind power. K. Wang et al. [11] clustered NWP data consisting of daily wind speed, pressure, humidity, and temperature by K-means and fed the data into a deep belief network for day-ahead prediction modeling, showing that reduced volatility and sophistication in NWP data drove the outperformance. This also revealed the difficulty of tuning hyperparameters in specific modeling problems, especially in network-based models.

Fortunately, decision tree algorithms do not require the adjustment of many parameters. S. Tasnim et al. [12] proposed a K-means cluster-based ensemble regression by linear and support vector regression for wind power forecast modeling and proved its superiority, with an up to 17.94% upgrade, by comparison with no-clustering and several ensemble models in seventy Australian wind sites. The upgrade, as compared with the baseline, is further enlarged to 20.63% by employing a transfer learning approach called multi-source domain adaptation, which includes a weighing method, innovatively calculated with data distributions by K-means clustering, to merge existing sites' information for new sites' power forecasting [13]. However, these efforts only used one certain clustering method and did not further explore other faster and more efficient clustering approaches.

As evidenced above, the effectiveness of cluster-based wind energy modeling analysis has been validated by multiple relevant models at wind sites worldwide, with engineering applicability and values. Nevertheless, except for the K-means algorithm, other well-suited clustering algorithms are rarely employed in this field. Interestingly, in this *journal*, Ref. [10] presented the significance of investigating different clustering approaches in wind power modeling. Wang et al. [14] conducted a self-organizing map clustering for classifying data and used neural networks and support vector machines as base learners to create a Bayesian model averaging ensembles for analyzing wind power. The model adapts to different meteorological conditions, but its clustering approach and learners are neural network-based and thus with high temporal complexity.

There remains a lack of comparative studies on ensemble learning wind power modeling with different clustering algorithms. Nor is there existing research that combines various clustering approaches-based stacking ensembles and considers the data diversities introduced by clustering for the modeling tasks. Both of these gaps are addressed in this study.

1.2. Contribution

Drawing on the literature review above, this study focuses on a wind farm in the Norwegian Arctic. A wind power modeling framework is proposed, which involves quantifying wind turbulence, clustering meteorological data, and ensemble learning. Firstly, an effective model integrating bagging and boosting is constructed. Secondly, four prominent clustering algorithms are systematically incorporated with models to form layered cluster-based ensembles and the best clustering approach is selected. Finally, stacking is employed to fuse these ensembles with different clusters to establish a more accurate model.

The principal contributions of this paper are thus as follows.

1. This paper experimentally proves that farthest first clustering is a distinctive approach in clustering wind data for power modeling compared to K-means, expectation-maximization, and Canopy clustering algorithms. The paper shows that even the worst-performing layered cluster-based ensemble outperforms the one without clustering. This indicates the similarities and dissimilarities in wind data. However, even though these data are not related to an individual wind turbine, they can still be significantly reflected in wind power in an implicit form.
2. Given the differences in results of different clustering algorithms, the paper proposes fusing layered ensembles with varying clusters with two-layer stacking to formalize a model that exceeds the optimal single clustering method. The stacking can more efficiently and quickly address the complex mapping task of nonlinear relationships between meteorological wind data and wind power.
3. The paper builds a procedure for determining the cluster number with a heuristic elbow chart, an empirical formula, and an X-means clustering approach. The procedure may be further developed and refined into a technique for identifying cluster numbers on other problems.
4. AdaBoost boosting with random forest bagging as its weak learner is apposite in the wind power models. These tree-based algorithms are computationally fast and parameter insensitive compared to the network-based ones. The proposed AdaBoost model statistically outperforms linear, neural network, and benchmark Adaboost approaches.

5. The quantization of wind turbulence intensities—both wind speed and direction that are rarely considered in related research—is applied to wind power modeling in a novel manner. The study finds that both intensities can serve as new features for considering wind volatility in the modeling.

The remainder of this paper is organized as follows. Wind meteorology and the use of data are described in Section 2. In Section 3, an elaborated description of the clustering approaches and statistical methods is presented. Section 4 shows the experimental procedure. Section 5 presents and discusses the obtained results. Finally, Section 6 concludes this work, noting its implications and outlook.

2. Wind power meteorology and data preparation

2.1. Wind power

Wind power generation is the conversion of wind kinetic energy into electricity. Ignoring losses in the conversion process, the actual output power of wind turbines can be expressed as in (1):

$$P = \begin{cases} 0 & v < v_{min} \\ \frac{1}{2} C_P(v)\rho A v^3 & v_{min} < v < v_n \\ P_n & v_n < v < v_{max} \\ 0 & v > v_{max} \end{cases} \quad (1),$$

where P represents the output power of the wind turbine (W); $C_P(v)$ represents wind energy utilization efficiency; ρ is the air density (kg/m^2); A represents the effective area swept by the wind turbine blades (m^2); v is the wind speed (m/s); $v_{\min}$, $v_{\max}$, and v_{n} respectively represent cut-in, cut-off, and rated wind speeds; and P_n is the rated wind power for the wind turbine. From (1), it is clear that the output of a wind turbine is primarily influenced by wind speed, air density, and swept area. Moreover, air density is primarily affected by temperature and pressure [15]. The swept area is related to the wind direction.

2.2. Quantification of turbulence in the wind

Turbulence arises when airflow moves through uneven landscapes or differences in air density. Turbulence is an immensely complicated flow phenomenon that is highly stochastic and difficult to characterize. In actual wind farm operations, turbulence is generated because of topographic and climate conditions and weak effects between wind turbines. Turbulence has a particularly strong impact on wind power production: given similar wind speed conditions, the higher the turbulence intensity, the higher the impact on wind farm output power [16]. The wind turbine's large inertia includes an impeller, whose rotation is behind wind speed change. Therefore, the turbine will not get the theoretically predicted wind force, and the power output will go down. Empirically, at low wind speeds, turbulence increases turbine power production. However, when wind speed approaches the turbine's furling speed, turbulence reduces production [17]. Nevertheless, turbulence is rarely considered in machine learning models of wind energy. An article in the *journal* [18] compared the effects of five popular learning algorithms and nine atmospheric variables on wind turbine power generation and found the following through statistical tests. First, for the five benchmark algorithms, the selection of atmospheric features for wind power modeling is more important; second, the top five features that are most influential for modeling are, in order, wind speed, turbulent kinetic energy, temperature, turbulence intensity, and wind direction. However, turbulent kinetic energy is seldom recorded by wind sites due to its measurement complexity. Therefore, turbulence intensity is considered as an input feature in this study.

Turbulence intensity, defined as wind speed standard deviation divided by the mean value over a short period [19], is the principal characteristic quantity of wind speed volatility. The turbulence intensity of direction is also applied as a quantitative tool to define turbulence behavior in wind direction. The turbulence intensities are shown in (2):

$$I_{SP} = \frac{S_{SP}}{SP},\ I_D = \frac{S_D}{D} \tag{2},$$

where I_{SP} and I_D are wind turbulence intensity of wind speed and direction; SP is wind speed; S_{sp} is its standard deviation of the previous ten minutes; D is wind direction index; S_D is its period standard deviation.

2.3. Data preparation

The study centers on a 54 MW wind farm designed in northern Norway, located about 500 km inside the Arctic Circle; it stands out as one of the largest wind farms in the Arctic. This farm's terrain features are a small hill, high steep mountains, and fjords, which are regarded as complex terrains. The wind power station company offered measurement of wind data with 10 min temporal resolution. We chose the five-dimensional meteorological wind data (wind speed and its variance, wind direction and its variance, and temperature) and power data from 0:00 1st January 2017 to 23:50 31st December 2017. Specifically, we calculated the sine values of wind direction and its standard deviation as indicators of wind direction and its fluctuations. Further, the turbulence intensities of wind speed and sine value of direction were computed as quantitative indices of wind turbulence. In summary, ten-minute resolution wind data—consisting of wind speed and sine direction and their turbulence intensities, temperature, and pressure—were employed to model wind power. These measurements contain small outliers and inevitable noise. However, the employed algorithm is insensitive to outliers, and this noise fits the standard normal distribution; therefore, they are not further considered. Because the scales of variables in the dataset vary widely, it is worth rescaling the original data into new data with similar proportions for each variable. Data normalization or standardization can increase model convergence speed and improve some algorithms' accuracy, especially in distance-based clustering [20].

3. Methodology

3.1. Chosen clustering approaches

Cluster analysis is an exploratory data mining technique for extracting useful information from high-dimensional datasets. It is a type of unsupervised learning approach to grouping similar hidden patterns [21] and classifying similar data into different subsets to give subset members identical attributes [22]. This classification requires quantifying the degree of similarity or dissimilarity between observations. The clustering results are strongly dependent on the kind of similarity metric used [23]. The cluster number is typically unknown and needs to be designated according to prior knowledge or determined by some method. Several clustering methods have been proposed. Ref. [24] offered some factors for choosing a clustering method. The method should be able to effectively and precisely find the suspected cluster types, and it must resist errors in the datasets; further, it must have the availability of computing power. This paper selects four clustering approaches: K-means, expectation-maximization, farthest first, and Canopy. The first one is a baseline method, and the other three can be regarded as competitors.

K-means: Among clustering algorithms, the K-means algorithm is one of the most popular and classical. It is a robust and versatile clustering algorithm proposed in [25],. The target of the K-means is to categorize observations into k clusters. K-means in this study is associated with Euclidean distance. Given a set of n data points $D = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ in $\mathbb{R}^d$and an integer k, the *K-means* problem is to determine a set of k centroids $C = \{\boldsymbol{c}_1, \dots, \boldsymbol{c}_K\}$ in $\mathbb{R}^d$ to minimize the following error function:

$$E(C) = \sum_{\mathbf{x} \in D} \min_{k=1,\dots,K} \|\mathbf{x} - \mathbf{c}_k\|^2 \qquad (3).$$

It is a combinatorial optimization that equals finding the partition of the n instances in k clusters whose associated set of mass centers minimizes Eq. (3) [26].

EM: The Expectation-Maximization (EM) algorithm is proposed by [27]. It provides a simple, easy-to-implement, and efficient tool for the learning parameters of a model [28], and is widely used. It finds the maximum likelihood or maximum posterior of the parameters in a probabilistic modeling process where the model relies on latent unobservable variables. The EM first initializes distribution parameters, then alternates between two steps: computing the expectation of variables based on the assumed initial parameters; and maximization, which gives a maximum likelihood estimate of the current parameters through the expectation values of the latent variables. The two steps repeat iteratively until the desired convergence is realized. When applied to clustering, the probabilistic model is established on the probability of each data sample to each cluster and distributes samples to the cluster with the largest possibility. The goal of EM clustering is to maximize the overall probability or likelihood of the clusters.

FF: The first utilization of Farthest First (FF) traversal is in [29]. FF is an effective greedy permutation method in computational geometry. Its underpinning is the traversal of a sequence of points in space where the initial point is specifically stochastic. The subsequent points are as remote as possible from the prior chosen set of points. FF clustering is the application of FF traversal in clustering, which was introduced in [30]. It is an optimized K-means with an analogous procedure, selecting the centroids first and assigning the samples to clusters with the maximum distance. Specifically, k numbers of centroids are generated by stochastically choosing a data point as the primary cluster centroid and greedily selecting the second centroid when it is FF from the first centroid. The process is repeated k times. As soon as all the centroids are recognized, FF assigns all the other data to the cluster in which the data have the nearest feature distance. In contrast to K-means, FF merely requires one traversal to cluster data. All the cluster centers are real data points, not geometric clustering centroids, and their position is fixed in the computation [31]. In most cases, the speed of clustering is considerably increased because fewer reassignments and adjustments are involved. The FF traversal is described in **Algorithm 1**.

Algorithm 1. Farthest First clustering Algorithm.

Farthest First Clustering (*D*: dataset, *k*: cluster number) {
select random data as the first point and first centroid;
// searching the data sample that is the farthest from the centroid
for (I=2,...,k) {
for (each remaining data sample in *D*) {
calculate the total distance to the existing centroids;}
select the sample with the largest distance as the new centroid;
label the centroids as {c_1, c_2, ..., c_k}}
//assignment the rest points {p_1, p_2, ..., p_n}
for (each point p_i) {
calculate the distance function dist to each fixed cluster centroid;
realize min {dist(p_i, c_1), dist (p_i, c_2), ..., dist (p_i, c_k)]}
put it to the cluster with minimum distance;} }

Canopy: Canopy clustering was introduced in 2000, and its central idea was to use a cheap, approximate distance measure to divide the data into subsets efficiently. This clustering decreases computing time over K-means and EM clustering methods by more than an order of magnitude and reduces errors on large datasets [32]. Unlike K-means, which only uses one distance, the Canopy algorithm uses two threshold distances, the larger loose distance T_1 and the smaller close distance T_2. It begins by removing a random point r sample from the original dataset and starting a *canopy* centered at *r*. It then approximates all distances between *r* and the remaining data r_i. If the distance is less than T_2, it places r_i in *r* canopy. If the distance is less than T_1, it removes r_i from a dataset. It repeats these steps until there is no more data to be clustered. However, Canopy needs to be tuned to the distance parameters and, according to [33], T_1 and T_2 can be obtained approximately using a heuristic based on attribute standard deviation.

3.2. Determining the cluster number

While various clustering methods are available, all of the mentioned methods require the cluster number before the clustering procedure. It is necessary to estimate the number due to the resulting partition of the data being dependent on its specification.

There have been energy studies using validated clustering methods, such as the elbow chart [10], Davies–Bouldin index [34], etc., to conduct data analysis. However, according to [35] and [36], there is no standard method for determining cluster numbers. Therefore, this paper combines three methods—formula, plotting, and information value—to comprehensively find a suitable cluster number for our meteorological wind data.

There is an empirical formula [37] to find the cluster number *k*. It is useful to check the range of *k* since it is not a precise approach.

$$k = 1 + 3.2\ log_{10}n \quad (4),$$

where *n* is the number of data points. This formula is inaccurate but can provide a reference for seeking *k*.

The elbow method is a visually heuristic technique for choosing cluster numbers [38]. The elbow principle's idea is that the total sum of squared errors between the sampling point in each cluster and the centroid (a smaller value means a more convergent result) is calculated with a series of *k* values. When the setup cluster number approximates the actual cluster number, the sum of squared errors will decrease swiftly. As the setup cluster number continues to grow, the Sum of Squared Errors (SSE) will continuously decrease, but more slowly [39]. Intuitive observation of turning points from elbow plots is sometimes vague. Still, it can provide a reasonable interval for searching for the value of *k*.

The X-means approach offers an effective tool for finding the exact *k*.

X-means: X-means is a variation of K-means clustering and can automatically determine the optimal cluster number in a dataset. It refines cluster assignment by repeatedly attempting subdivision segments and keeping the best resulting splits. It searches the space of cluster locations and the cluster number to optimize the Bayesian Information Criterion (BIC) measure [40]. The main parameters for X-means are the lower and upper bounds of the cluster number, which are found in the above two methods. It includes two steps that are repeated until they reach the required convergence. Primarily, the K-means algorithm is utilized to cluster the given dataset. Each cluster centroid is divided into two parts in opposite directions along a stochastic vector. The K-means algorithm is locally operated within the old cluster and generates two new clusters. By comparing the BIC scores of the original clustering structure with a new one, the splitting is either made or not. The idea is that splitting a single cluster into two clusters increases the BIC score, with two clusters being more probable than one. When k reaches the set upper bound, the splitting stops, and the algorithm reports BIC scores for each k value.

3.3. Modeling ensemble learning algorithm

The fundamental idea behind ensemble learning is to ensemble multiple algorithms or models to achieve an integrated model with better predictive performance [41]. The ensemble method can tactfully partition the dataset into smaller ones, train them separately, and then combine them with some strategies. The main strategies can be categorized into three groups: boosting, bootstrap aggregating (shortened as "bagging"), and stacking.

In the bagging procedure, new training sets are formed by taking from the original training set with a put-back. The averaging method for each new result of the training set is applied to reach the final result in a regression. Random forest [42] is an efficient bagging algorithm that uses decision trees as its base learners and offers decent performance and low computing costs. It is an improvement in the decision tree algorithm in which, essentially, multiple decision trees are merged. The creation of each tree depends on an independent bagging subset. Each tree in the forest has the same probability distribution. The final regression value can be determined by averaging each predictive value from each tree. Since random forest introduces perturbations in sampling and features, it dramatically improves generalization and avoids overfitting. Further, it can handle high-dimensional data without feature selection, and crucial features are derived during the training process [43].

Boosting [44] is an approach that boosts weak learners to strong learners. Adaptive boosting (shortened as AdaBoost) is a representative boosting algorithm. It continually builds weak learners to emphasize (with larger weights) on samples mislearned in the prior learner until the number of learners reaches the setup value or the loss function reaches a threshold.

For the regression problem, the weighted average is used to eventually obtain predicted values. Adaboost is highly accurate, can adequately construct weak learners, and is not susceptible to overfitting. Meanwhile, it is sensitive to anomalous samples (which may receive large weights in iterations), affecting the performance of strong learners. For the numeric output for the strong learner $h_i(\mathbf{x}) \in \mathbb{R}$, weighted averaging (5) is used for the final result [45].

$$H(\boldsymbol{x}) = \sum_{i=1}^{M} w_i h_i(\boldsymbol{x}) \quad (5),$$

where w_i is the weight of a weak learner and $w_i \geq 0, \sum_{i=1}^{T} w_i = 1$.

Stacking is a representation learning technique that can extract valid features from data by employing meta-learning algorithms to learn how to optimally combine predictions from many base learners. Several different base models are first trained with the original dataset. A new model named meta-learner is then trained with each of the previous models' outputs to get a final output [46]. The stacking result is typically better than its single base learner since the fusional ensemble combines varying types of base learners. The applied stacking is shown in **Algorithm 2** [47].

Algorithm 2. Stacking algorithm with four base learners and one meta-learner.

Input: Dataset $\boldsymbol{D} = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_m, y_m)\}$

Base learner varying clustering approaches-based Adaboost algorithms $\mathfrak{L}_1, \ldots, \mathfrak{L}_4$;

Meta-learner linear regression $\mathfrak{L}$

Process:

for $t = 1,2,3,4$ **do**

$h_t = \mathfrak{L}_t(D)$;

// Train base leaners by $\mathfrak{L}_t$

end for

// Generate training set for meta-learner

$\boldsymbol{D}' = \boldsymbol{\emptyset}$;

for $i = 1,2, \ldots, m$ **do**

for $t = 1,2,3,4$ **do**

$z_{it} = h_t(\boldsymbol{x}_i)$;

end for

$\boldsymbol{D}' = \boldsymbol{D}' \cup \left((z_{i1}, z_{i2}, z_{i3}, z_{i4}), y_i\right)$

end for

// Meta-learner h' is established

$h' = \mathfrak{L}(D')$

Output: $H(\boldsymbol{x}) = h'(h_1(\boldsymbol{x}), h_2(\boldsymbol{x}), \ldots, h_T(\boldsymbol{x}))$

A two-layer assemblage structure for regression, which can be categorized as a kind of layered cluster-based or oriented ensemble named by [48], is adopted to optimally incorporate clustering results generated separately by the four above clustering approaches into the AdaBoost mechanism. The ensemble structure achieves excellent learning ability and prediction accuracy by mapping the first-layer clustering to the second-layer ensemble regression [21].

3.4. Proposed modeling strategy

A proposed framework for clustering approach comparisons is displayed in Fig. 1. It is inspired by wind energy meteorology, clustering approaches, and ensemble learning. The framework is a two-layer architecture, with four clustering algorithms in layer 1 and AdaBoost in layer 2. Specifically, the random forest is the weak learner for the AdaBoost, and Reduced-Error Pruning TREE (REPTREE) [49] is introduced to replace the decision tree in the random forest to reduce overfitting that may be caused by the complicated ensemble model structure.

Take K-means clustering as an example. Layer 1 uses Section 3.2 to identify the cluster number k and clusters the wind data into k clusters. Layer 2 employs each cluster to train the AdaBoost to learn AdaBoost and establish k submodels with labels. Subsequently, the test data, one by one, are classified into an existing cluster and loaded into the trained AdaBoost submodel corresponding to the cluster for wind power modeling, and overall performance is calculated.

Analogously, the above procedure is also applied to EM, FF, and Canopy clustering approaches. More experimental details are presented in Section 4.

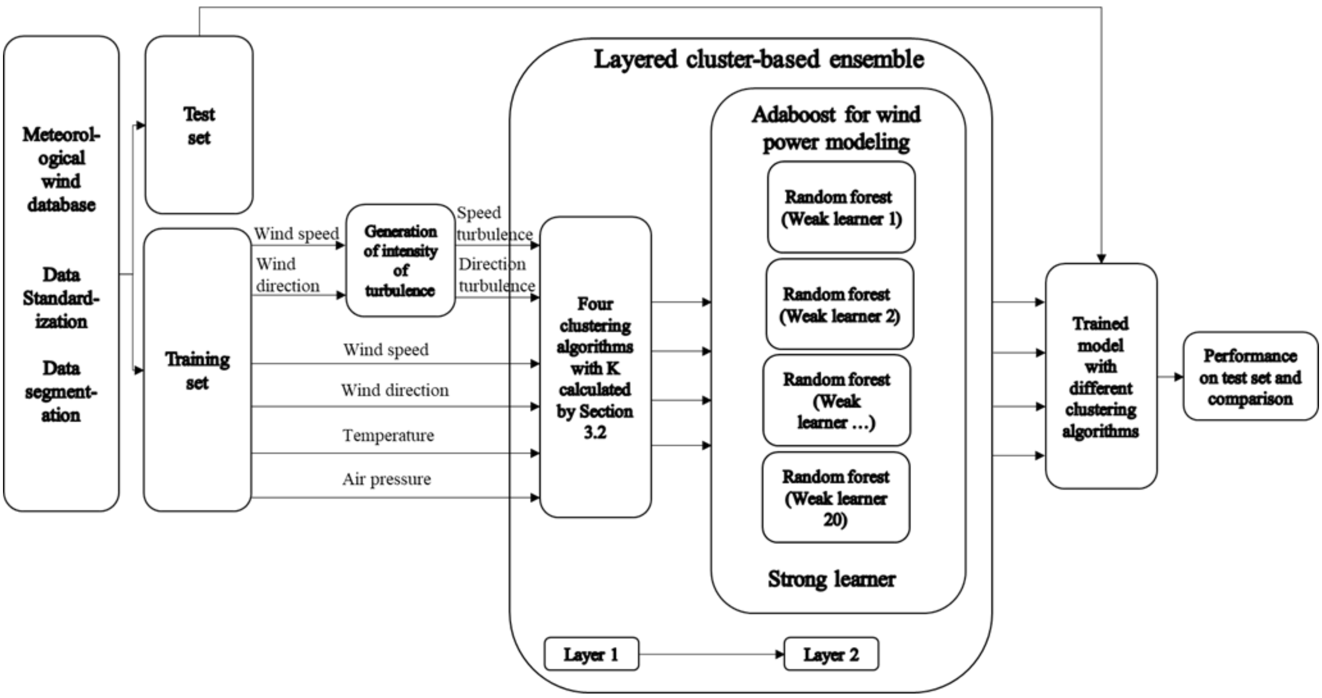


Fig. 1. The procedure of the proposed strategy for wind power modeling.

Regarding stacking ensemble modeling, a novel method is put forward. It also consists of two layers, the first being base learners (AdaBoost models with the four different clustering algorithms) and the second being linear regression Eq. (5) with Tikhonov regularization $\lambda \parallel w \parallel_1$ (also named ridge regression [49] Eq. (6) to avoid overfitting caused by the complex model structure [50]). The reasons for this configuration are the following. First, the first layer has diversity in the layered ensembles based on four clustering algorithms and may deeply extract data features and transmit them to the second layer. Second, the major risk of the second layer is that it learns the generated data from the first layer and is vulnerable to overfitting, so linear regression with a regular term is the learning algorithm in this layer. The first layer procedure is the same as in Fig. 1. It generates four sets of simulated power on training and test sets. Subsequently, the second layer uses the measured power and four generated power sets as the dependent and independent variables, respectively, to build ridge regression on the training set and employs the learned regression to predict the power with the simulated test power on the test sets.

$$f(\boldsymbol{x}) = \boldsymbol{w}^{\top}\boldsymbol{x} + b \qquad (6),$$

with a loss function $J = \frac{1}{n}\sum_{i=1}^{n} (f(\boldsymbol{x}_i) - y_i)^2 + \lambda \parallel w \parallel_1$:

$$\begin{array}{ll} \underset{w,b}{min} & \frac{1}{n}\sum_{i=1}^{n} (\boldsymbol{w}^{\top}\boldsymbol{x}_i + b - y_i)^2 \\ \text{s.t.} & \parallel w \parallel_1 \leq t \end{array} \qquad (7).$$

3.5. Model evaluation metrics and multiple comparisons

Two metrics are utilized in evaluating the performance of different models in the test set. The first one is Normalized Mean Absolute Error (NMAE), while the second is Normalized Root Mean Square Error (NRMSE). They are negatively oriented, which means the smaller value is related to better performance. The NRMSE assigns a higher weight to larger errors because of the square calculation, meaning it punishes substantial prediction errors and reveals whether the regression has noticeable error variance.

$$NMAE = \frac{\sum_{i=1}^{n}|prediction_i - observation_i|}{n} / \frac{\sum_{i=1}^{n} observation_i}{n} \qquad (8),$$

$$NRMSE = \sqrt{\frac{\sum_{i=1}^{n}(prediction_i - observation_i)^2}{n}} / \frac{\sum_{i=1}^{n} observation_i}{n} \qquad (9).$$

Two statistical approaches are used to check whether there are statistically significant differences between the model's performance. The Friedman test is used to check for differences in performance across multiple trials [51]. It tests column effects after adjusting for possible row effects.

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

H_a: They have a significant difference.

Its statistic F is shown as:

$$F = \frac{12n}{k(k+1)}\left[\sum_{i=1}^{k} r_i^2 - \frac{k(k+1)^2}{4}\right] \quad (10),$$

where k is the number of columns, r_i is the mean value of row i, which follows $\chi^2_{(k-1)}$ under H_o.

Furthermore, the Tukey method is used for computing confidence intervals between the means of two populations. It is expressed as follows:

$$(\bar{Y}_1 - \bar{Y}_2) \pm \frac{q_{k,n-k,1-\alpha}}{\sqrt{2}} \cdot \sqrt{MSE} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (11),$$

where q is the Gaussian q-distribution, k is the number of populations, and n is its total size; MSE is the Mean Square Error within groups.

4. Experiment setup

This study extracts meteorological wind data from the Norwegian Water Resources and Energy Directorate, including a few abnormal negative values, at which the wind farm did not generate electricity but consumed grid power. All weather data are normalized as inputs to the models. First, the wind data are divided into a training set, accounting for 90%, and a test set, accounting for 10%. To fully apply the data, avoid overfitting, and improve generalization in modeling [52], 10-fold cross-validation is used in the training. Then, weather data are harnessed in the test set to calculate the corresponding wind power, which is compared to the actual power data to obtain performance metrics.

For the benchmark model, the processed training data are directly fed into the AdaBoost with random forest (Layer 2 in Fig. 1) (AdaRF). The number of iterations is set to 100 (trade-off between performance and computing speed). Random forest is the AdaBoost inner weak learners, and the number of REPTREE in each random forest is set to 10. The competitors are linear regression (LR), artificial three-layer neural networks (16 nodes in the hidden layer, which is found by a grid search from 6 to 20) (ANN), and AdaBoost with 20 decision trees (achieved by a grid search from 5 to 30 with an interval of 5) as its weak learners (AdaDT).

Regarding the ensemble model based on clustering approaches, the range of cluster numbers for the weather data is first found by the elbow graph and empirical formula (4). Its exact value is determined using the X-means clustering method. Then, the four aforementioned clustering approaches are used to group the data in the training set and categorize the test data into established clusters to find the best-performing clustering algorithm. Finally, stacking is employed to combine layered cluster-based ensembles with different clustering algorithms to further explore avenues to upgrade power modeling.

In this study, wind power modeling is f realizing the relationship between wind power and wind weather $\boldsymbol{W_t}$. The model is shown in (12).

$$\hat{P_t} = f_t(\boldsymbol{W}_t) + e \quad (12),$$

in which

$$\boldsymbol{W} = [V, IV_{turbulance}, sin(\theta), Isin_{turbulance}(\theta), T, P] \quad (13),$$

where $\hat{P_t}$ is modeling wind power; f_t (.) is the model that needs to be implicitly realized; $\boldsymbol{W}_t$ represents weather data that will be clustered by the four clustering approaches; e is the model error.

5. Experiments and results

5.1. Feature ranking and comparison for modeling without clustering

The training wind data are initially harnessed to establish a multivariate linear wind power regression model to check the feature attributing degree. The diagnosis (T statistic and its corresponding two-tailed p-value [53]) for interpretation of each feature is shown in Table. 1.

Table 1. The wind features were selected by the statistical diagnosis of linear regression.

Futures	***V***	***$IV_{turbulance}$***	***Sin(θ)***	***$Isin_{turbulance(\theta)}$***	***T***	***P***
T statistic; *p*-value	250.79; <0.0001	7.84; <0.0001	-21.77; <0.0001	3.02; 0.0025	23.68; <0.0001	0.67; 0.5040

Note: the term is shown as "*T* statistic; *p*-value." The H_0 is where the interpretation equals zero and its H_a is where the term is not zero; when the p-value is smaller than the set confidence level of 0.05, the H_0 is rejected and the feature is attributed to the linear model.

All meteorological features are statistically significant in the linear modeling except pressure P. The features' importance may be approximatively ranked by absolute values of T statistics in a descending scale as V, *Sin(θ)*, T, $IV_{turbulance}$, $Isin_{turbulance(\theta)}$, P. Although pressure does not contribute to the linear regression, all above meteorological features are still accounted for in the modeling as pressure values are relatively stable and the presented models are clustering and tree models demanding low computations and feature selection.

To enhance the verifiability of modeling results, the year is split into four quarters, Q1, Q2, Q3, and Q4, for individual power modeling. The statistical variability among quarterly data is initially analyzed in Table 2. Statistics and distribution disparities between meteorological wind and power quarterly datasets can be summarized, and quarterly data differ from yearly data. Therefore, separate modeling on these datasets can strengthen the proposed strategy's credibility.

Table 2. The statistics of the yearly and quarterly wind data.

Statistics / Dataset	**Average**	**Standard deviation**	**Skewness**	**Kurtosis**
Year	<0.0001	0.9983	4.9493	100.7258

Q1	0.1988	0.8981	3.9305	87.7615
Q2	-0.0283	0.8729	5.4827	129.5278
Q3	-0.0356	0.8421	5.4604	118.6208
Q4	-0.1350	0.8873	4.7543	113.6719
CoV for Statistics	-53584	0.0586	0.1157	0.1316

Note: The different variables are standardized to similar scales, so the statistics of the various variables in the dataset are averaged and shown. Coefficient of Variation (CoV) is defined as the ratio of standard deviation to mean.

The four quarterly and yearly normalized training data are separately entered into the proposed AdaRF, benchmarking LR, ANN, and AdaDT to map the relationship between wind data and wind power. Fig. 2 shows the results. Both NMAE and NRMSE increase significantly as time grows. The NMAE and NRMSE of AdaRF are significantly lower than the results obtained from multivariate linear regression. The average NMAE and NRMSE decrease by 52.98% and 46.31%, respectively. The AdaRF decrease in NMAE and NRMSE (corresponding to the model improvement) is also evident when compared with ANN (NMAE 19.54% and NRMSE 10.43%) and AdaDT (NMAE 29.31% and NRMSE 17.95%). Fig. 3 displays the modeling power of a day, from which AdaRF appears close to real values but with several errors in points. This means the proposed AdaRF enables accurate power modeling based on weather data but still leaves room for refinement.

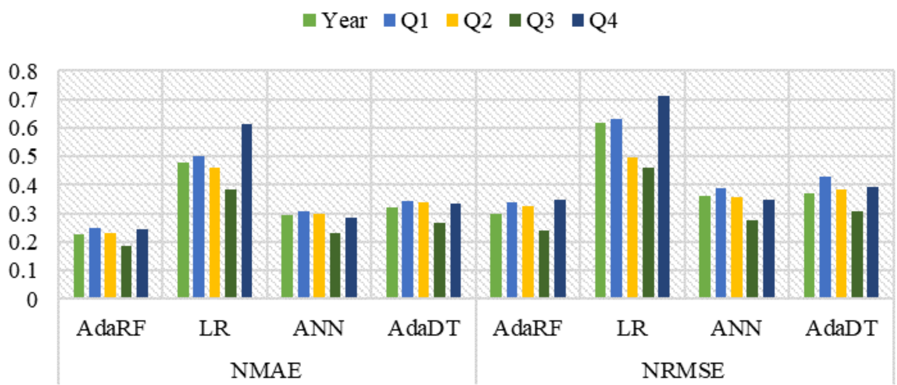


Fig. 2. The performance comparison for power modeling without clustering.

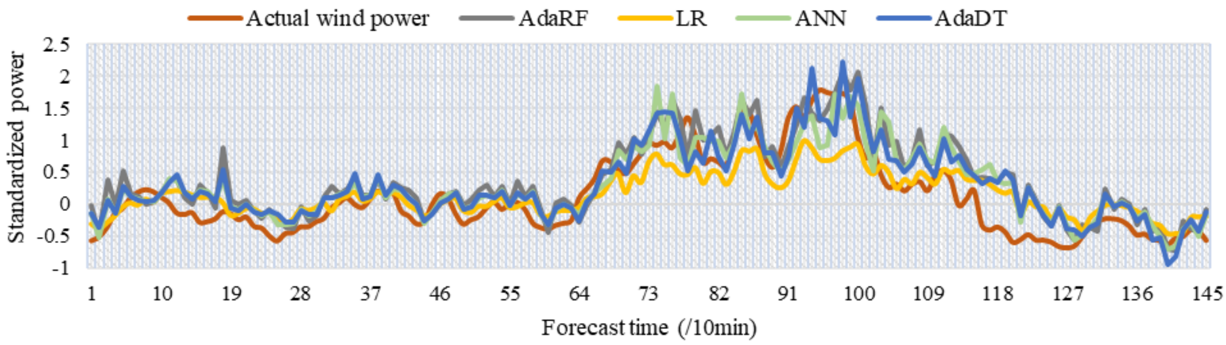


Fig. 3. The wind power modeling from weather data in the wind farm.

5.2. Determination of an appropriate cluster number

The size of the yearly dataset is 52,560; *k* is calculated to be approximately 16 in (4). Selecting this value as the midpoint, the total Sum of Squared Errors (SSE) of K-means is calculated with a starting point of *k* equals 2 and an endpoint of *k* equals 30. The elbow plot is drawn and displayed in Fig. 4. The precise value of the elbow point for the total sum of squared errors can be determined only approximately since the process is by intuition and experience. However, Fig. 4 still shows an interval, the cluster number $k \in [10,20]$, in which the decline of SSE begins to flatten from steep, and where the elbow point belongs to.

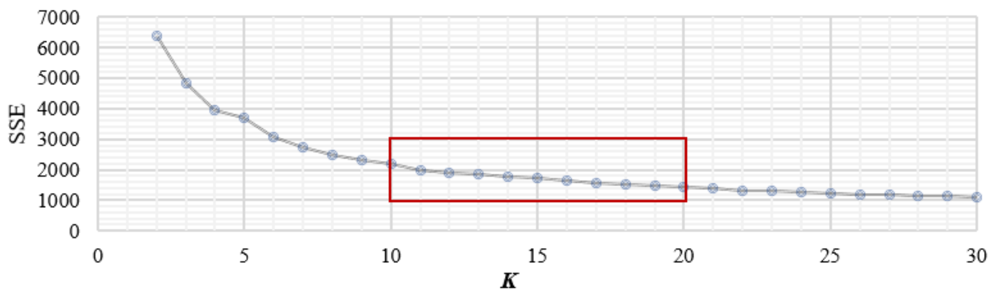


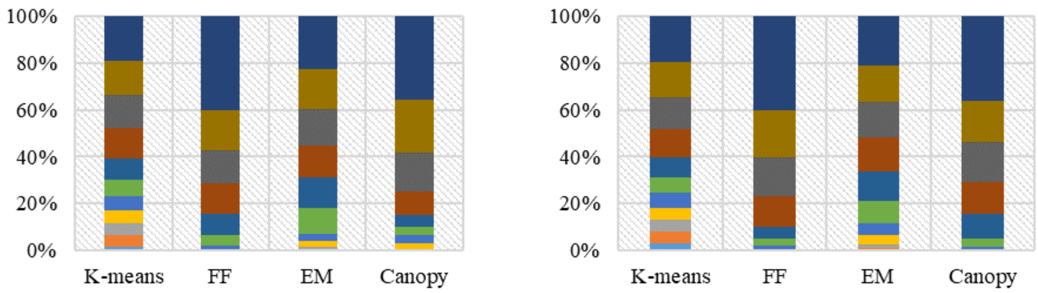
Fig. 4. The elbow plot for finding cluster number *k*.

To demonstratively find the precise value of *k*, an X-means approach is adopted. The lower and upper bounds of *k* are set as 10 and 20 respectively according to the interval formerly found in Fig. 4. The optimized BIC score is 69,956.78 with a proper *k* value for the meteorological wind data equaling 11.

Analogously, the *k* values for the four quarterly wind data, Q1, Q2, Q3, and Q4, are decided as 14, 8, 9, and 12, respectively.

5.3. Comparison of different clustering approaches in modeling

For the yearly dataset, the four clustering approaches yield varying numbers of samples per cluster, albeit at the same cluster number. Based on the four separate clustering methods separately with 11 clusters, four complete layered cluster-based ensembles are developed for clustering comparisons. Firstly, the wind weather data with different clustering approaches in the training and test sets are shown in Fig. 5. Each color represents a cluster, and the vertical axis shows the percentage of each number in the total dataset.



(a) Training set (b) Test set

Fig. 5. The clusters number percentage of different clustering approaches.

The various clustering methods produce wildly different clustering results, even with the same *k*. The cluster sample number variance analysis reveals that the K-means method produces more homogeneous clustering than other methods. Even single-digit sample percentages are seen in the FF and Canopy algorithms for the training set. Apart from the K-means, all of the other three algorithms generate clusters that exceed one-fifth of the sample size. Second, the four clustering methods' yearly meteorological wind test data are loaded into the layered cluster-based ensemble for wind power modeling. The NMAE and NRMSE are displayed in Fig. 6.

The model based on FF clustering intuitively presents the smallest NMAE and NRMSE, 32.35% and 33.64% reduction without clustering; the second smallest is the model with Canopy. The models with these two clustering approaches significantly improve their performance compared to the ones without clustering, while the K-means and EM algorithms also upgrade their models' abilities.

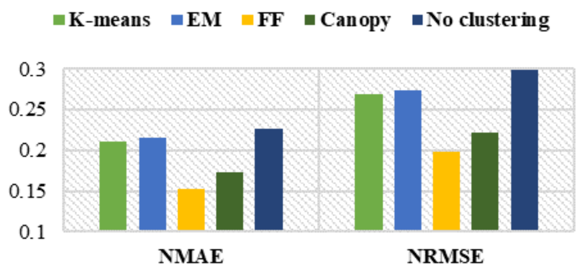
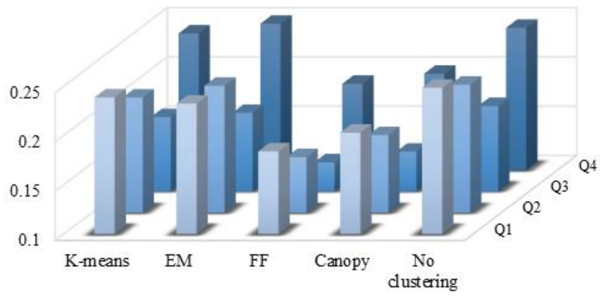
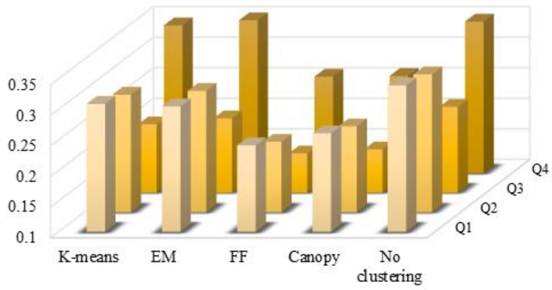


Fig. 6. The NMAE and NRMSE of the power modes with different clustering methods for the yearly dataset.

To further elaborate comparisons of clustering methods, analogously, layered cluster-based ensemble modeling with different clustering approaches is conducted on four quarterly wind datasets, and their NMAE and NRMSE are displayed in Fig. 7. The ranking of the models built on quarterly data is the same as those built on yearly data. Strengths in the FF and Canopy clustering algorithms are evident in each quarter. Moreover, a result is derived that the 3rd quarter-power model performs the best, followed by the 2nd quarter. This illustrates a more clear relationship between wind data and power from April to September, which is consistent with the intuition that the area has milder weather during this period.



(a) NMAE



(b) NRMSE

Fig. 7. The NMAE and NRMSE of the power models with different clustering methods for quarterly data.

Multiple comparisons are conducted between the metrics from different cluster-based models between quarters. The Friedman test p-values of NMAE and NRMSE are both 0.0056 and much smaller than the confidence level of 0.05, so the null hypotheses are rejected. This leads to the conclusion that there are differences between the metrics of the various ensembles.

Table 4 compares the average NMAE and NRMSE for models with different kinds of clustering against the ones without. Quarterly evenly, the new model reduces NMAE and NRMSE by 13.94% and 17.45%. Furthermore, performance improvement between the two models generally slumps from summer to winter.

Table 4. The average performance improvement between the models with clustering and the corresponding one without.

	Q1	**Q2**	**Q3**	**Q4**
NMAE	13.66%	15.06%	16.29%	10.74%
NRMSE	17.73%	19.60%	20.08%	12.37%

Regarding the best-performing FF clustering of the yearly dataset, table 5 compares NMAE and NRMSE for the FF-based model to the original model. Both NMAE and NRMSE have an approximately 23% to 34% decrease, which indicates this clustering approach is twice as good as the average clustering method in our case. Further, the superiority of FF is different with the quarter: the model boosting results are more noticeable during warm periods compared to those during cold seasons.

Table 5. The performance improvement between the models with FF clustering and the baseline.

	Q1	Q2	Q3	Q4
NMAE	25.98%	31.94%	30.60%	23.24%
NRMSE	28.83%	33.85%	31.68%	25.88%

Further, the Canopy clustering-based approach also displays a satisfactory result, which improves the modeling performance by averages of over 20% in NMAE and 25% in NRMSE, respectively. EM and K-means clustering-based approaches have relatively similar performance. However, they are still not as good as the FF and Canopy, as the above yearly analysis shows.

Collectively, the Tukey method calculates the intervals with 95% confidence of metrics difference; Table 6 shows the bounds of these intervals between the yearly and quarterly models with no clustering and ones with varying clustering algorithms. The upper and lower bounds of the metrics difference between no and FF clustering are positive, indicating that the superiority of the FF is statistically significant across multiple datasets. Moreover, the upper bounds of all other differences are greater than the lower bounds of absolute values, illustrating that normally distributed differences have positive means, which describes the other cluster-based models as outperforming no clustering in a probabilistic sense. Therefore, the edges of wind data clustering, ranking as FF, Canopy, K-means, and EM in order, before the layered ensemble modeling procedure, are demonstrated in our datasets.

Table 6. The bounds for paired comparisons of clustering across yearly and quarterly datasets.

No clustering	v.s.	K-means	EM	FF	Canopy
NMAE	Lower Bound	-0.0363	-0.0414	0.0176	0.0011
	Upper Bound	0.0585	0.0534	0.1124	0.0959
NRMSE	Lower Bound	-0.0572	-0.0642	0.0122	-0.0030
	Upper Bound	0.1082	0.1012	0.1776	0.1624

5.4. Stacking ensemble power modeling

Section 5.1 demonstrates that the proposed AdaRF outperforms three other benchmarks (ANN, AdaDT, and LR in descending order). Section 5.3 illustrates in Fig. 5 the four clustering algorithms yielding highly diverse clusters, and the cluster-based models work better. The AdaRF model is further refined by implementing a two-layer stacking structure (Cls-AdaRF): The first layer takes the four clustering outcomes in Fig. 5 as inputs to AdaRF to generate four-layered cluster-based ensembles; the second layer combines these ensembles outputs by linear regression to yield final simulations. Its performance is compared with that of AdaRF without clustering (NCl-AdaRF) in Section 5.1 and AdaRF with FF clustering (FF-AdaRF) in Section 5.3. The comparison in Fig. 8 shows that Cls-AdaRF decreases more NMAE and NRMSE in percentage than NCl-AdaRF (NMAE 31.89% and NRMSE 34.74%) and FF-AdaRF (NMAE

4.32% and NRMSE 5.71%). Its edge over FF-AdaRF indicates that stacking combined with four different clustering algorithms outperforms the best-layered ensemble with a single clustering. These model quarterly variations are similar to those in Section 5.3.

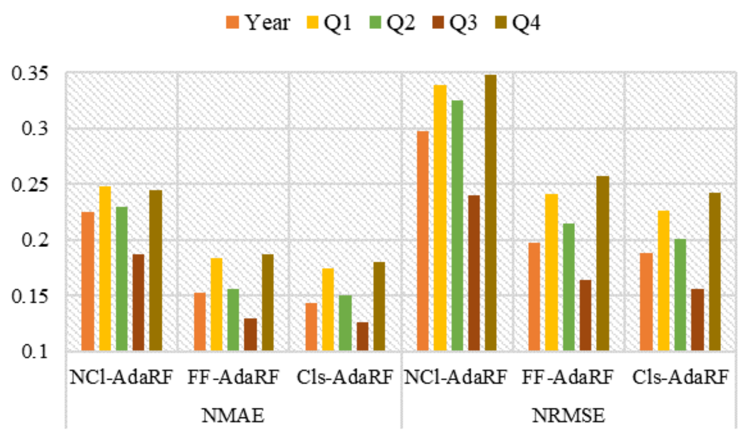


Fig. 8. The NMAE and NRMSE of the power models with different clustering methods for quarterly data.

Likewise, placing benchmark algorithms ANN, AdaDT, and LR into this process eventually yields three cluster-based stacking ensembles, denoted as Cls-LR, Cls-ANN, and Cls-AdaDT. These models, including those that are run on the datasets separately, and their NMAE and NRMSE, are compared to those of Cls-AdaRF. Table 7 shows the difference intervals calculated by the Tukey method and reveals that, except for Cls-AdaRF vs. Cls-ANN in NMAE (where the upper bound is considerably close to zero), all the intervals are negative, indicating a 95% statistical significance among datasets for the Cls-AdaRF model's strength.

Table 7. The bounds for paired comparisons of stacking across yearly and quarterly datasets.

Cls-AdaRF	**v.s.**	**Cls-LR**	**Cls-ANN**	**Cls-AdaDT**
NMAE	Lower Bound	-0.3735	-0.1398	-0.1654
	Upper Bound	-0.2154	0.0183	-0.0073
NRMSE	Lower Bound	-0.3896	-0.1835	-0.2011
	Upper Bound	-0.2097	-0.0036	-0.0212

The percentage reductions in NMAE and NRMSE for Cls-AdaRF versus other models (corresponding to model improvement) are further calculated and presented in Fig. 9. On average, Cls-AdaRF delivers over 20% improvements over its three competitors (within a standard deviation; Cls-AdaRF outperforms Cls-LR, Cls-ANN, and Cls-AdaDT by about 60%, 25%, and 35%, respectively.).

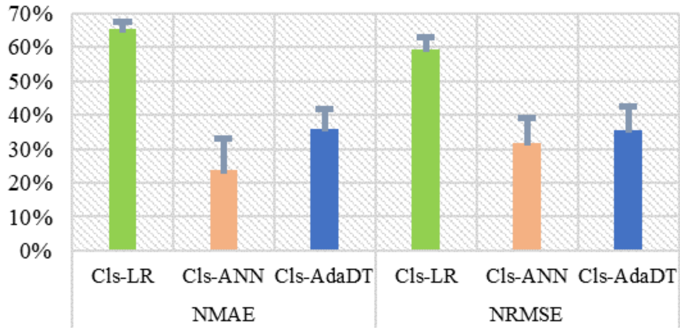


Fig. 9. The NMAE and NRMSE improvement of Cls-AdaRF versus other models for the yearly dataset.

Altogether, the Cls-AdaRF model is inferred to be a superior wind power model because it not only outperforms AdaRF without clustering but is also better than other stacking models.

6. Conclusions

This paper presents an ensemble learning approach that combines bagging, boosting, and stacking for modeling wind power from meteorological data. To mine the inherent characteristics of the data, four clustering approaches are used to process inputs for the layered ensembles. Then, the layered cluster-based ensembles are fused within the stacking framework. The proposed models' superiority is verified by diverse comparisons.

The AdaRF can accurately model wind power. The algorithm circumvents issues of an equal weighting of each tree in RF and AdaBoost and allows each learner to boost incrementally, and eventually creates a model with a good generalization. The overall performance of the proposed method is on average 33.94% in NMAE and 24.90% in NRMSE, lower as compared to the benchmarks in the cases excluding clustering.

As no standard methods for identifying the cluster number exist, the paper uses a heuristic elbow graph, an empirical formula, and the X-means clustering algorithm to precisely determine the implied number for meteorological data. Interestingly, the number for the yearly dataset is 11, which is close to the month's number. This result suggests that there may be analogous phenomena to measured wind data with monthly periodicity.

A comparative study of AdaRF based on different clustering methods reveals, firstly, that the model with clusters significantly performs better than the model without, regardless of what clustering approach is employed. This suggests that similarities within the wind power data can correspond to similarities within the weather data. Secondly, among these clustering methods, the model with FF clustering provides the best modeling results. The reason is that FF is built on finding the data point furthest from the previous centroid as the new one; in other words, it emphasizes large differences between clusters. Upon this clustering, the fluctuations among the original meteorological data are considerably diminished, which in turn corresponds to a smoother wind power output and increases the accuracy of the wind power model. The fast computability and accuracy of FF also suggest that the clustering technique can be applied to ultra-short-term wind power models. Thirdly, Canopy is the fastest among the four clustering methods and achieves comparable results. Therefore, Canopy can also serve as a favorable clustering approach when wind weather datasets are considerably large.

Finally, the wind power model is further strengthened by using stacking Cls-AdaRF to fuse the layered ensembles with four clustering approaches. It can be interpreted as Cls-AdaRF working as a representation learning—that is, effective features are automatically collected from raw data and fed into the second layer via multiple learners in the first layer; the second layer compiles and aggregates these features through linear regression with a regular term and effectively outputs simulations.

Practically, given that the proposed well-performing wind power model does not involve complex network training and extensive parameter tuning and wind energy physical modeling, it can be easily transferred to other energy utilization scenarios.

Further research, as suggested by the above conclusions, is needed to deeply optimize the base learners of stacking and their combination algorithms to deliver faster and more accurate modeling. Another direction is to incorporate this article's in-the-now power modeling approach and meteorological data with historical wind power to achieve efficacious short-term power forecasting.

References

[1] J. Lee *et al.*, "Global wind report 2019," *Brussels: Global Wind Energy Council (GWEC),* 2020.

[2] Z. Tian, Y. Ren, and G. Wang, "Short-term wind power prediction based on empirical mode decomposition and improved extreme learning machine," *Journal of Electrical Engineering & Technology,* vol. 13, no. 5, pp. 1841-1851, 2018.

[3] A. Marvuglia and A. Messineo, "Monitoring of wind farms' power curves using machine learning techniques," *Applied Energy,* vol. 98, pp. 574-583, 2012.

[4] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renewable energy,* vol. 37, no. 1, pp. 1-8, 2012.

[5] R. Liu, M. Peng, and X. Xiao, "Ultra-short-term wind power prediction based on multivariate phase space reconstruction and multivariate linear regression," *Energies,* vol. 11, no. 10, p. 2763, 2018.

[6] J. Ma, M. Yang, X. Han, and Z. Li, "Ultra-short-term wind generation forecast based on multivariate empirical dynamic modeling," *IEEE Transactions on Industry Applications,* vol. 54, no. 2, pp. 1029-1038, 2017.
[7] D. Niu, D. Pu, and S. Dai, "Ultra-short-term wind-power forecasting based on the weighted random forest optimized by the niche immune lion algorithm," *Energies,* vol. 11, no. 5, p. 1098, 2018.
[8] Y. Dong, H. Zhang, C. Wang, and X. Zhou, "Wind power forecasting based on stacking ensemble model, decomposition and intelligent optimization algorithm," *Neurocomputing,* vol. 462, pp. 169-184, 2021.
[9] V. Kushwah, R. Wadhvani, and A. K. Kushwah, "Trend-based time series data clustering for wind speed forecasting," *Wind Engineering,* p. 0309524X20941180, 2020.
[10] L. Dong, L. Wang, S. F. Khahro, S. Gao, and X. Liao, "Wind power day-ahead prediction with cluster analysis of NWP," *Renewable and Sustainable Energy Reviews,* vol. 60, pp. 1206-1212, 2016.
[11] K. Wang, X. Qi, H. Liu, and J. Song, "Deep belief network based k-means cluster approach for short-term wind power forecasting," *Energy,* vol. 165, pp. 840-852, 2018.
[12] S. Tasnim, A. Rahman, A. M. T. Oo, and M. E. Haque, "Wind power prediction using cluster based ensemble regression," *International Journal of Computational Intelligence and Applications,* vol. 16, no. 04, p. 1750026, 2017.
[13] S. Tasnim, A. Rahman, A. M. T. Oo, and M. E. Haque, "Wind power prediction in new stations based on knowledge of existing Stations: A cluster based multi source domain adaptation approach," *Knowledge-Based Systems,* vol. 145, pp. 15-24, 2018.
[14] G. Wang, R. Jia, J. Liu, and H. Zhang, "A hybrid wind power forecasting approach based on Bayesian model averaging and ensemble learning," *Renewable Energy,* vol. 145, pp. 2426-2434, 2020.
[15] C. Jung and D. Schindler, "The role of air density in wind energy assessment–A case study from Germany," *Energy,* vol. 171, pp. 385-392, 2019.
[16] K. Kaiser, W. Langreder, H. Hohlen, and J. Højstrup, "Turbulence correction for power curves," in *Wind Energy*: Springer, pp. 159-162, 2007,
[17] W. D. Lubitz, "Impact of ambient turbulence on performance of a small wind turbine," *Renewable Energy,* vol. 61, pp. 69-73, 2014.
[18] M. Optis and J. Perr-Sauer, "The importance of atmospheric turbulence and stability in machine-learning models of wind farm power production," *Renewable and Sustainable Energy Reviews,* vol. 112, pp. 27-41, 2019.
[19] M. Türk and S. Emeis, "The dependence of offshore turbulence intensity on wind speed," *Journal of Wind Engineering and Industrial Aerodynamics,* vol. 98, no. 8-9, pp. 466-471, 2010.
[20] A. Ng, "Advice for applying machine learning," in *Machine learning*, pp. 3-15, 2011.
[21] B. Verma and A. Rahman, "Cluster-oriented ensemble classifier: Impact of multicluster characterization on ensemble classifier learning," *IEEE Transactions on knowledge and data engineering,* vol. 24, no. 4, pp. 605-618, 2011.
[22] P. Berkhin, "A survey of clustering data mining techniques," in *Grouping multidimensional data*: Springer, pp. 25-71, 2006.

[23] A. Serra and R. Tagliaferri, "Unsupervised learning: clustering," *Encyclopedia of Bioinformatics and Computational Biology; Elsevier: Amsterdam, The Netherlands,* pp. 350-357, 2019.
[24] G. W. Milligan, "Clustering validation: results and implications for applied analyses," in *Clustering and classification*: World Scientific, pp. 341-375, 1996.
[25] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, no. 14: Oakland, CA, USA, pp. 281-297, 1967.
[26] M. Capó, A. Pérez, and J. A. Lozano, "An efficient approximation to the K-means clustering for massive data," *Knowledge-Based Systems,* vol. 117, pp. 56-69, 2017.
[27] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society: Series B (Methodological),* vol. 39, no. 1, pp. 1-22, 1977.
[28] C. B. Do and S. Batzoglou, "What is the expectation maximization algorithm?," *Nature biotechnology,* vol. 26, no. 8, pp. 897-899, 2008.
[29] D. J. Rosenkrantz, R. E. Stearns, and I. Lewis, Philip M, "An analysis of several heuristics for the traveling salesman problem," *SIAM journal on computing,* vol. 6, no. 3, pp. 563-581, 1977.
[30] D. S. Hochbaum and D. B. Shmoys, "A best possible heuristic for the k-center problem," *Mathematics of operations research,* vol. 10, no. 2, pp. 180-184, 1985.
[31] R. D. H. Devi, A. Bai, and N. Nagarajan, "A novel hybrid approach for diagnosing diabetes mellitus using farthest first and support vector machine algorithms," *Obesity Medicine,* vol. 17, p. 100152, 2020.
[32] A. McCallum, K. Nigam, and L. H. Ungar, "Efficient clustering of high-dimensional data sets with application to reference matching," in *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2000, pp. 169-178.
[33] R. R. Bouckaert *et al.*, "WEKA manual for version 3-9-1," *University of Waikato, Hamilton, New Zealand,* pp. 47-49 2016.
[34] Y. Li, B. Wang, Z. Yang, J. Li, and C. Chen, "Hierarchical stochastic scheduling of multi-community integrated energy systems in uncertain environments via stackelberg game," *Applied Energy,* vol. 308, p. 118392, 2022.
[35] T. M. Kodinariya and P. R. Makwana, "Review on determining number of Cluster in K-Means Clustering," *International Journal,* vol. 1, no. 6, pp. 90-95, 2013.
[36] K. P. Sinaga and M.-S. Yang, "Unsupervised K-means clustering algorithm," *IEEE access,* vol. 8, pp. 80716-80727, 2020.
[37] V. Braverman, A. Meyerson, R. Ostrovsky, A. Roytman, M. Shindler, and B. Tagiku, "Streaming k-means on well-clusterable data," in Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms: SIAM, pp. 26-40 , 2011.
[38] A. Ng, "Clustering with the k-means algorithm," *Machine Learning Course,* 2012.
[39] C. Yuan and H. Yang, "Research on K-value selection method of K-means clustering algorithm," *J—Multidisciplinary Scientific Journal,* vol. 2, no. 2, pp. 226-235, 2019.
[40] D. Pelleg and A. W. Moore, "X-means: Extending k-means with efficient estimation of the number of clusters," in *Icml*, vol. 1, pp. 727-734, 2000,.

[41] L. Rokach, "Ensemble-based classifiers," *Artificial intelligence review,* vol. 33, no. 1-2, pp. 1-39, 2010.
[42] L. Breiman, "Random forests," *Machine learning,* vol. 45, no. 1, pp. 5-32, 2001.
[43] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?," *The journal of machine learning research,* vol. 15, no. 1, pp. 3133-3181, 2014.
[44] L. Breiman, "Bias, variance, and arcing classifiers," Tech. Rep. 460, Statistics Department, University of California, Berkeley, 1996.
[45] A. Bertoni, P. Campadelli, and M. Parodi, "A boosting algorithm for regression," in *International conference on artificial neural networks*, Springer, pp. 343-348, 1997.
[46] L. Breiman, "Stacked regressions," *Machine learning,* vol. 24, no. 1, pp. 49-64, 1996.
[47] Z.-H. Zhou, "Ensemble learning," in *Machine Learning*: Springer, pp. 181-210, 2021.
[48] A. Rahman and B. Verma, "Novel layered clustering-based approach for generating ensemble of classifiers," *IEEE Transactions on Neural Networks,* vol. 22, no. 5, pp. 781-792, 2011.
[49] G. C. McDonald, "Ridge regression," *Wiley Interdisciplinary Reviews: Computational Statistics,* vol. 1, no. 1, pp. 93-100, 2009.
[50] S. Džeroski and B. Ženko, "Is combining classifiers with stacking better than selecting the best one?," *Machine learning,* vol. 54, no. 3, pp. 255-273, 2004.
[51] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference: Revised and Expanded*. CRC press, pp. 353-393, 2014.
[52] S. Yadav and S. Shukla, "Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification," in *2016 IEEE 6th International conference on advanced computing (IACC)*, IEEE, pp. 78-83, 2016:
[53] G. A. Seber and A. J. Lee, *Linear regression analysis*. John Wiley & Sons, pp. 265-327, 2012.

Paper III. A

Comparative study of data-driven short-term wind power forecasting approaches for the Norwegian Arctic region

https://doi.org/10.1063/5.0038429

Comparative study of data-driven short-term wind power forecasting approaches for the Norwegian Arctic region (Postprint)

(This is a postprint version, the print version copyrights belong to AIP Journal of Renewable and Sustainable Energy)

Hao Chen[1], Yngve Birkelund[1], Stian Normann Anfinsen[2], and Fuqing Yuan[1]

[1] Department of Technology and Safety, [2]Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø 9019, Norway

Abstract

This paper conducts a systemic comparative study on univariate and multivariate wind power forecasting for five wind farms inside the Arctic area. The development of wind power in the Arctic can help reduce greenhouse gas emissions in this environmentally fragile region. In practice, wind power forecasting is essential to maintain the grid balance and optimize electricity generation. This study firstly applies various learning methods for wind power forecasting. It comprehensively compares the performance of models categorized by whether considering weather factors in the Arctic. Nine different representative types of machine learning algorithms make several univariate time series forecasting, and their performance is evaluated. It is demonstrated that machine learning approaches have an insignificant advantage over the persistence method in the univariate situation. With numerical weather prediction wind data and wind power data as inputs, the multivariate forecasting models are established and made one hour to six hours in advance predictions. The multivariate models, especially with the advanced learning algorithms, show their edge over the univariate model based on the same algorithm. Although weather data are mesoscale, they can contribute to improving the wind power forecasting accuracy. Moreover, these results are generally valid for the five wind farms, proving the models' effectiveness and universality in this regional wind power utilization. Additionally, there is no clear evidence that predictive model performance is related to wind farms' topographic complexity.

Key words: wind energy, machine learning, power forecasting, numerical weather prediction, Arctic

1 INTRODUCTION

To prevent global average temperatures from rising 1.5°C above pre-industrial level, the renewable energy percentage must increase from 20% to 67% of global energy production from 2018 to 2040 [1]. Wind energy is one of the fastest-growing renewable energy sources. It is considered an attractive alternative to conventional electricity sources generated from fossil fuels. Wind power is extensive, and its capacity has surged from 9,936 MW in 1998 to 564,347 MW in 2018, with an annual growth rate of 22.4% in the last 20 years [2]. Along with the electricity grid adds wind power penetrations, the unstable grid factors are also increased, which are undesirable to the power system practical and safe operations. So, it is crucial to use proper methods to understand wind power production and harness proper methods to make forecasts of the electricity generated by the wind parks.

Norway has a cold climate and a 25,148km coastline, both of which are generally characterized by an abundance of wind energy resources, and it is with a complex terrain consisting of mountains, valleys, and fjords, making the wind change dramatically and unpredictably.

Wind power prediction can be divided into ultra-short-term prediction, short-term prediction, medium-term prediction, and long-term prediction [3]. Ultra-short-term forecasts are predictions made from few minutes to 30 minutes in advance; the short-term are forecasts made from 30 minutes to 48 hours ahead, the medium-term refers to predictions made days, weeks, or months earlier, and the long-term is made years in advance.

In wind engineering, hourly wind power forecasting is an essential part of the short-term prediction, whose main applications are maintaining real-time grid operations and keeping operational security in the electricity market [4].

In this study, five wind parks in the Norwegian Arctic regions are taken as the target. Table 1 serves as a summarized comparison in terms of installed capacity, location, and site ruggedness (RIX) [5] of the five sites.

1.1 Related work

In literature, there is much research on wind power forecasting using multiple analysis methods from many perspectives. A preliminary study on wind energy forecasting considered the use of statistical methods. Still, there is a trend of using machine learning algorithms for the forecast. Machine learning is an emerging artificial intelligence approach that attempts to provide learning capabilities for computers or other equipment without clear operations [6]. It aims to develop strategies and algorithms that learn patterns from training data and make predictions. It can be an alternative tool in wind engineering, apart from the statistics and physical methods [7] to forecast wind power with historical wind data. In particular, deep learning, which has emerged in recent years, offers a promise of automating pattern recognition and solving problems such as complex wind power predictions. However, there is a well-known rule called the No Free Lunch (NFL) theorem in the context of supervised machine learning, which states that averaged over all optimization problems, all non-resampling optimization algorithms perform equally well [8]. Due to geographical and engineering reasons, the most suitable machine learning algorithms for wind power prediction for different wind farms vary.

Ref. [9] did a systematic literature review and found that artificial neural networks are used more frequently to predict wind energy and provide better results than other methods, as demonstrated with more than 180 references in the five years. Specifically, Ref. [10] focused on a wind farm in north Iran at 5-min time interval predictions and found that the adaptive neuro-fuzzy inference system outperforms the other five data mining algorithms: random forests, M5Rules, k-nearest neighbor, support vector machine, and multilayer perceptron. Ref. [11], based on the Portuguese wind power data throughout 2010–2014, also showed that the adaptive neural fuzzy inference system was the best performer. The artificial neural networks and the radial basis function network RBFN-OLS also delivered strong performances. Ref. [12] demonstrated that the proposed hybrid artificial neural network is effective and efficient for wind power forecasting in a Danish dataset. Ref. [13] used an approach combining the infinite feature selection with the recurrent neural networks and proved its edge in a dataset from the National Renewable Energy Laboratory. Ref. [14] investigated five years of wind observation data of Nigde, Turkey, and found that eXtreme gradient boost, support vector regression, and random forests algorithms are powerful in forecasting long-term daily total wind power and the absolute shrinkage selector operator is the worst algorithm due to its linear basis. Ref. [15] mixed basic Multi-Layer Perceptron to complex deep learning neural networks to conduct the power prediction of a wind farm located in the Ecuadorian mountains. The hybrid model is shown to be more advantageous than a single model.

Notably, this journal emphasizes the basis and the state-of-the-art of wind power forecasting. Ref. [16] offered a detailed adaptabilities analysis of the support vector machine, genetic algorithm backpropagation, and radial basis function for wind power forecasting based on three wind farms in China. Ref. [17] noted the essence of deep learning in predictions; sometimes, the forecasts did not need models based on truly deep neural networks, but they offer a sound workflow for correctly developing a proper forecast model.

Meanwhile, much research concerns a particular class of machine learning algorithms, such as kernel methods and neural networks-based methods. There are a few studies that make comparisons between types of algorithms. The reason is applying machine learning always needs tuning the hyperparameters, which makes the comparison rather sophisticated. However, choosing an algorithm less sensitive to parameters or using a suitable method of adjusting parameters, and scale-up and diversification data set can help deal with the problem. Ref. [18] proposed a two-stage wind power forecast method with meteorological factor and fault time and compared the method performance with support vector machines, artificial neural network, generalized regression neural network, and radial basis function and found the edge of the first algorithm. Moreover, there is a lack of complete comparative research on both univariate and multivariate time series forecasting with data science for wind energy prediction in the Arctic region characterized by dense air and excellent wind resources.

1.2 Objective and contributions

The objective and main contributions of this study can be summarized as follows:

1. The paper does a systematic study of the time series forecast for five wind parks generating power with sufficient wind potentials in the Norwegian Arctic region. We mainly focus on investigating the multivariate wind power forecasting models by considering Numerical Weather Prediction (NWP) data.

2. For brief experimental univariate power forecasts. The persistence model and nine machine learning benchmarking algorithms are researched in forecasting models and compared their performance from an algorithm perspective. We find the persistence model performs almost equally to machine learning models in our cases. The result also proves conclusions from Ref.[19]; those classical methods may dominate univariate time series forecasting. However, we find that its performance drops more quickly with the forecast time step rises. Considering the contingency of parameters tuning and computational complexity of the learning algorithm, it is suggested that statistical modeling methods should be primarily considered in the forecasting.

3. The multivariate models with mesoscale NWP wind data, although the data resolution scale is larger than the wind parks area, as inputs can slightly gain prediction accuracy compared with the univariate models with the same algorithm. Moreover, the multivariate models reduce performance slower than the univariate models, which indicates the informative complementary role that the weather data play in the model

4. These five wind farms have different complex terrains and climates. The wind park with complex terrains implies that the NWP wind results are not as accurate as their counterparts of plain landscapes. However, there is no significant evidence that prediction results are related to the ruggedness index of wind parks from our results.

2 DATA DESCRIPTION AND PREPARATION

2.1 Wind power locations

Northern Norway has a complex terrain consisting of fjords, mountains, and valleys that goes from the coast into a moderately high inland along the border to northern Sweden and Finland. The terrain elevation around each wind park is also shown in Fig.1, and their coordinates and heights are listed in Table 1. Nygårdsfjellet wind park is located in a valley, far from the open sea, that reached approximately 450 meters elevation. The mountains south and north of the valley limit the main wind direction to be west-east, and high wind events are expected during the winter season [20]. Havøygavlen, Kjøllefjord, and Fakken wind parks are located close to the open sea and on relative flat hills where large nearby fjords affect both wind direction and speed. Raggovidda wind park is also located near the open sea but on a flat mountain that does not have any vegetation. This location is well known for adequate wind resources and produced power with relatively high capacity factors (the ratio between the real and designed power production) over several years.

2.2 Norwegian Meteorological Institute (MET Norway) numerical weather prediction

The Scandinavian weather institutions use a weather model for weather forecasts named MEPS (Ensemble Prediction System). The weather model makes ensemble forecasts, starting from a composition of several forecasts and quantifying the outcome space of possible weather developments, which depends on the weather itself rather than looking at a single estimate [21]. The NWP model is a complex mathematical model of the atmosphere that divides the earth surface into grids [22]. The spatial resolution of the grid determines how to simulate meteorological processes with different accuracy levels, limiting the quality of forecasts.

A study conducted by MET Norway has demonstrated that the regional NWP models with higher resolution did not result in better wind power forecasts for some Norwegian wind farms [23]. Therefore, in this study, we use the NWP data with 2.5km horizontal resolution, which is regarded as a relatively coarser resolution in wind forecasting.

2.3 Data description and scaling

The hourly power data of five wind farms, measured hourly, used in the research is provided by the Norwegian Water Resources and Energy Directorate (NVE). We choose the wind power data from 0:00 1st January 2017 to 23:00 31st December 2017; the measured data are 8,760 for each wind farm. The total number of wind power data is 43,800. The location, annual mean powers, the standard deviation, and the capacity factor of the five wind farms in 2017 are also shown in Table 1.

Table 1. The location and statistics of power data

Wind Park	Location °N / °E	Height [m]	RIX	Designed power [MW]	Mean power [MW]	Standard deviation [MW]	Capacity factor
Nygårdsfjellet	68.504 / 17.879	410	0-5	32.2	11.132	11.833	34.57%
Fakken	70.098 / 20.081	95	5-10	54.0	15.239	15.858	28.22%
Raggovidda	70.769 / 29.094	440	0-5	45.0	21.782	16.869	48.40%
Kjøllefjord	70.922 / 27.268	280	10-20	39.1	12.349	12.786	31.58%
Havøygavlen	71.012 / 24.589	220	5-10	40.5	10.311	11.037	25.46%

The NWP wind forecast data are extracted from MET Norway operational weather forecast model MEPS and considering that forecasts need two hours to be calculated as usual. The forecasts are all initiated at 00, 06, 12, and 18 UTC.

Wind power generating is mainly affected by wind speed, wind direction, and air density, which is impacted by temperature and air pressure [24]. In this study, we use the variables acquired at time *t*; such as measured generating wind power, NWP wind speed, NWP wind direction (radian system), NWP surface air pressure, and NWP 2 meters above the ground temperature to predict the wind power generated at *t*+*n*, where *n*, ranging from 1 to 6, is the time delay in hours. The NWP wind data are summarized and shown in Table 2. All the items show variables with the *mean value* (*standard deviation*) form, and the relative air pressure means the real local air pressure minus the standard atmospheric pressure (101,325Pa).

Table 2. The statistics of the original NWP data

Wind park	Speed (m/s)	Direction (radical system)	Temperature (°C)	Relative air pressure (Pa)
Nygardsfjellet	8.096 (5.038)	-0.065 (0.431)	0.045 (7.441)	-5795.564 (1246.119)
Fakken	6.948 (3.885)	0.151 (1.032)	4.193 (5.109)	-1091.373 (1284.892)
Raggovidda	9.49 (5.101)	0.011 (0.855)	-0.91 (6.256)	-5148.793 (1277.989)
Kjøllefjord	7.9 (4.213)	0.15 (0.962)	1.23 (5.763)	-2848.796 (1292.669)
Havoygavlen	8.335 (4.434)	0.136 (0.872)	2.953 (5.33)	-1750.36 (1309.583)

Data scaling is a standard approach to normalize data. An important reason for data scaling is that the algorithm converges faster with feature scaling than without it [25]. And it is convenient to compare the model performance with similar data scales. The wind power data is scaled with min-max normalization between 0.2 and 0.8.

$$x' = a + \frac{(x - min(x))(b-a)}{max(x) - min(x)} \qquad (1)$$

where a and b are the minimum and maximum values of the normalization scale.

2.4 Stationary test

The power data for five wind farms can be treated as five univariate time series sequences. Time series can be divided into stationary and non–stationary data sequences. Whether or not a time series is stationary has long been a question of major interest in the field of time series analysis [26]. Statistical regression processes can analyze stationary time series; meanwhile, the non–stationary time series change their statistical properties with time [27]. So, the forecasting for non–stationary time series is more problematic than for stationary ones. Augmented Dickey-Fuller test (ADF) is a widely used method for testing the null hypothesis that a unit root is present in a time series sample. Its principle is to check whether a unit root is present in a sequence. If no unit root presents, the sequence is stationary; otherwise, it is nonstationary. The ADF test is a standard method to test the stationarity of economic time series [28]The ADF test utilizes the autoregressive process and optimizes its parameters for various lag values. A null hypothesis test can conduct the ADF test application in testing the stationary of a time series sequence.

H_0: the time series is nonstationary, which means it shows a time-dependent structure.

H_a: the time series is stationary.

The ADF test is conducted on the five wind farms power data. The results show that all the null hypotheses are rejected with critical values that are much lower than 5%, demonstrating the five-time series power data are stationary. They also show the power data sequences do not have trends and seasonality, which means there is no need to divide the annual wind power data into monthly or seasonal sequences in forecasts.

3 FORECASTING ALGORITHMS FOR WIND POWER

For each utilized forecasting model, numerous changes are proposed by researchers [29], and it is impossible to conduct all of the existing differences in models. Therefore, our strategy is to consider each benchmarking model for different algorithms for hourly wind power forecasting. The ten prediction models, one baseline statistical model, and nine machine learning models are chosen because they are commonly used models.

1. Persistence Model (PE)
2. Support Vector Regression (SVR)
3. K-Nearest Neighbor regression (KNN)
4. MultiLayer Perceptron (MLP)
5. Radial Basis Functions (RBF)
6. Classification and Regression Trees (CART)
7. Random Forest (RF)

8. Stochastic Gradient Boosting (SGB)
9. SVR optimized with Genetic Algorithm (GA-SVR)
10. Long Short-Term Memory neural networks (LSTM)

Method 1 is the reference in the performance comparison, whereas approaches 2 to 6 are representative and widely used machine learning algorithms, and 7 and 8 are two representative types of ensemble machine learning techniques. There are two main types of ensemble learning methods: bootstrap aggregating (bagging) and boosting [30]. The rest two are emerging trends of predictive algorithms, representing hybrid methods and deep learning, respectively. The following offers a brief description of each algorithm. Due to the page limitation, more detailed descriptions of these algorithms are available in the references of this article.

The PE model takes that the power at time $t + n$ equals t the power at t, n is the next n steps in time series. It assumes that the atmospheric conditions change stationarily.

SVR is a regression model provided by the support vector machine algorithm, which tries to identify the hyperplane that maximizes the margin between two classes and minimizes the total error under tolerance [31]. SVR conducts a penalty with C (complexity penalization term) and achieves the best trade-off between the empirical error and the model complexity. SVR can perform a nonlinear regression because it provides kernel functions (like linear, polynomial, and Gaussian) that map data from the input space to a high dimensional feature space in which regression is conducted. The value of C is taken from a validation test for $C \in \{0.01,0.1,1,10,100\}$. C is found with a value of 1 corresponding to the best performance with the Gaussian kernel function.

KNN regression focuses on feature similarity determined by distance functions, like Euclidean, Manhattan, or Minkowski distance, measurements for data samples [32]. The K parameter, which implies the input consists of the k closest training sample subsets, determines the performance of the algorithm; a large K value can reduce the noise in the regression process, but it also leads to a risk of overfitting. In the study, we conduct a grid search for K from 1 to 10 in experiments to find an appropriate K value.

MLP is a network of simple neurons named perceptrons. The perceptron forms a linear combination based on its input weights and calculates the output through a nonlinear activation function [33]. MLP is a versatile approach for forecasting; it can find nonlinear structures in a problem and model a linear regression process. MLP is a parameterized model. We can manage the MLP complexity by choosing the number of hidden nodes and the type of activation functions. Specifically, the sigmoid function is usually used as the activation function in MLP regression problems. In the study, the topology of the MLP consists of three layers; namely, the number of nodes for the input layer equals input numbers, a hidden layer with ten nodes, and an output layer with one node.

RBF networks are feedforward and similar in structure to the MLP. The radial basis functions are harnessed as their activation functions. Their output is a linear combination of radial basis functions (radially symmetric around a center) applied to values of inputs neurons. The RBF also has a fast and efficient training process of both linear and nonlinear mappings. In the study, the RBF topology is the same as the MLP model.

CART regression is based on a tree-like recursive partition of inputs [34]. The CART is made of internal decision nodes and end leaves. Given a test data set, the terminal leaves are decided by different training sample properties. Besides, a series of tests are created and utilized in the decision nodes, which can define where the inputs should be classified to specific nodes whose splitting will most significantly reduce the mean square error. Moreover, a final decision tree is realized when the mean square error is smaller than a threshold.

Bagging is a unique variant of the model averaging approach to reduce the prediction variances by repeatedly creating subsets of original data to train the machine learning model. RF is an efficient bagging ensemble algorithm and delivers sound capability and low computational cost. RF is based on the establishment of a multitude of sub learners. Each learner is trained by using a bootstrap sample extracted from the whole training set. The forest of learners produces ensemble regression values. The final result is determined, e.g., by averaging over the ensemble [35]. RF has only one difference from the general bagging decision tree: it uses an improved decision tree algorithm, selecting a random subset of features at each sample selection in the training process.

Boosting is an iterative technique that uses the last classification to adjust the weights of nodes constantly in the learning process. SGB is a trendy and widely used boosting learning algorithm. It constructs regressions by sequentially fitting a base learner to current "pseudo"-residuals by least-squares in each iteration. It can improve the accuracy of gradient boosting and training speed by incorporating randomization into the learning procedure [36].

Genetic Algorithm (GA) is one of the well-regarded evolutionary algorithms. It mimics the Darwinian theory of survival of the fittest and arrives at such configuration via cycles consisting of individual population generation, selection, crossover, and mutation phase [37]. During the process, the population of candidates originates from a combination of the offspring and survivors of the previous generation or a randomly generated configuration. The population then faces two selection phases that decide which candidates do not survive into the next generation and then decide which candidates may produce child candidates. This filtering uses a fitness function. GA-SVR is using the GA in optimizing complexity penalization term C of SVR in the training process and has become a so-called hybrid forecasting method.

Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture that is mainly used in the field of deep learning [38]. Unlike standard feedforward neural networks, LSTM has feedback connections, and unique LSTM units consist of input, output, and forget gates. It can process not only single data points but also entire sequences of data (such as speeches or videos). It can also be used in time series forecasting with proper pre-treatment of data. We utilize the Vanilla LSTM that is with one hidden layer, including LSTM units and one output to predict.

4 EXPERIMENTAL SETUP

Multi-steps wind predictions are required in short-term wind energy generation. In the study, we make direct forecasting, which only uses actual measured values as model inputs. It builds n different prediction models for n steps ahead forecasting. The benefit of direct forecasting is that it does not use previous prediction values to forecast the values with higher steps, which means the prediction is not affected by the cumulative error in the forecasting process.

Wind power prediction from t_1 to t_3 for the univariate forecasting with one-dimensional input wind power at time t_0. t_1 to t_6 is for the multivariate forecasting with five-dimensional input (wind power, NWP wind speed, NWP wind direction, NWP temperature, and NWP air pressure at time t_0) because according to Ref. [7], the weather factors are recommended considered after 3 hours. Namely, we conduct two modeling processes for each wind park. The multivariate model is displayed in equation 2.

$$\hat{P}_{t+n} = f_{t+n}(P_t, W_t) + e_n, \ n = 1,2,3,4,5,6 \quad (2)$$

where $\hat{P}_{t+n}$ is the n steps wind power forecasting, f_{t+n} is the forecasting model, P_t and W_t represent the wind power and NWP weather data at time t, e_n is the model error.

The source data set is divided into 66% for training the model and 34% for testing the models' performance and carrying out comparisons.

4.1 Cross-validation and evaluation Metrics

One of the most critical and popular validation methods for machine learning is k-folds cross-validation; it is more suitable for relatively small and limited data set compared with the train and test split validation approach because it can ensure that there is a chance for every sample in the original data to appear in the training as well as the validation process. In this research, we use k = 10 in the implementation of the whole training set. According to a study by Kohavi, this is usually a pretty good choice [39].

Two metrics are used in evaluating the performance of the different kinds of algorithms for wind power forecasting. The first metric is the Mean Absolute Error (MAE); the second metric is the Root Mean Square Error (RMSE). The definitions for MAE and RMSE are shown in equations 3 and 4. Both are negatively oriented metrics, meaning the lower scores are related to better performance [40]. In our cases, if an approach has a low MAE but a high RMSE, it generally predicts smoothly and efficiently but has a higher population of large significant forecasting errors that are weighted significantly by RMSE.

$$MAE = \frac{\sum_{i=1}^{n} |prediction_i - observation_i|}{n} \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (prediction_i - observation_i)^2}{n}} \quad (4)$$

4.2 Forecasting procedure

For the procedure of univariate hourly wind power forecasting, the wind power data are normalized and tested with ADF stationary tests. Then we use the PE model and nine machine learning algorithms to do three steps forecasts.

Moreover, to further improve wind power forecasting accuracy and make full use of the NWP data, we established a multivariate forecasting model based on SVR, MLP, RF, and LSTM algorithms, shorten as '*NWP* plus the abbreviation', based on the univariate forecasting results and some recommendations for publications [41, 42]. The performance is compared with their counterparts in univariate cases. The procedure of forecasting is illustrated in Fig. 1.

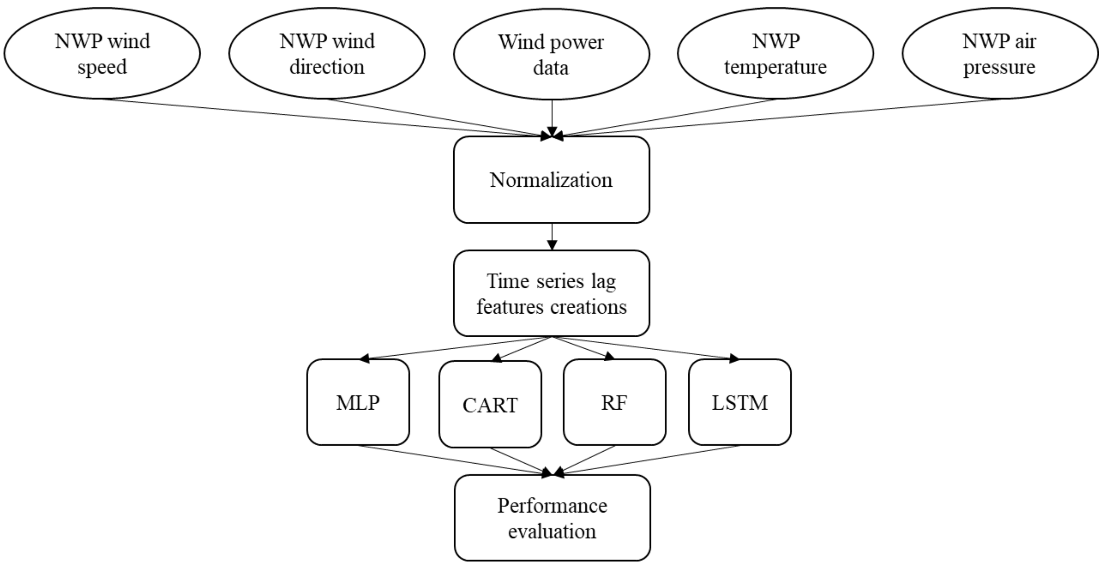


Fig. 1. Procedure for the multivariate hourly wind power forecasting

5 RESULTS

We conduct a univariate wind power forecasting with the aforementioned ten algorithms and adds the NWP wind data as inputs to create new MLP, CART, RF, and LSTM multivariate forecasting models for five wind parks.

For the univariate forecasting, the performance comparison of ten models is briefly conducted. For multivariate forecasting, the four multivariate models' performance is compared with their counterparts in univariate forecasting cases.

5.1 Univariate forecasting

Regarding MAE results, the MAEs of all models increase as the forecasting step adds. The average MAE of three steps predictions is displayed in Fig. 3. The PE models perform similarly with SVR and GA-SVR models in all wind parks. The KNN, MLP, RBF, CART, and LSTM have similar MAEs, which are more unsatisfactory results than PE and SVR and GA-SVR have. The ensemble learning methods show the highest MAEs.

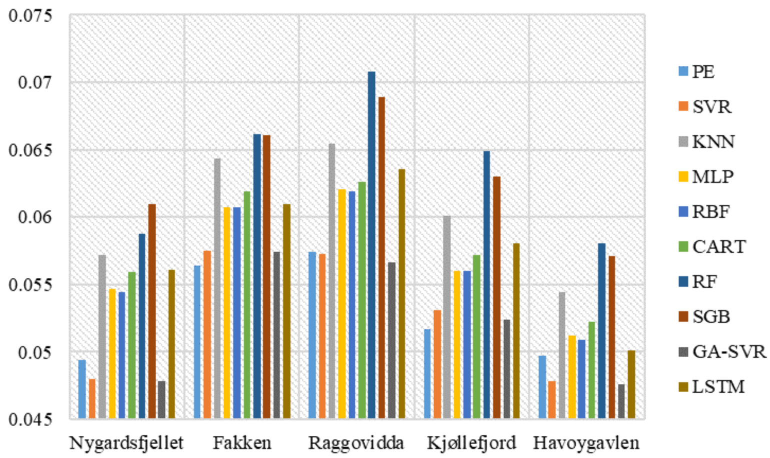


Fig. 2. The average MAE of ten forecasting models for five wind farms

RMSEs of all models have positive correlations with the forecasting step from 1 to 3. The main RMSE of three steps forecasts is shown in Fig. 4. The MLP, RBF, and GA-SVR have the best performance in RMSE, in which MLP has the lowest overall RMSE. The PE, KNN, CART, SGB, and LSTM have similar RMSEs, and nearly all of them have more inferior results than MLP, RBF and GA-SVR do. The RF model has the highest RMSE as it does in MAE analysis.

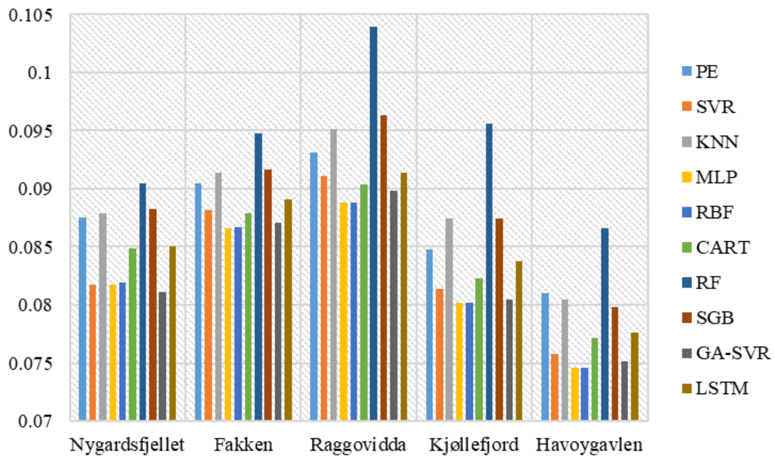


Fig. 3. The average RMSE of ten forecasting models for five wind farms

5.2 Multivariate forecasting

The MAE of the multivariate forecast for the five wind parks is displayed in Fig. 5. We predict one hour to six hours of wind power. As the forecasting step increases, the MAE of all models increases, and the rising speed gets slower. For each forecasting step, the NWP machine learning models have lower MAE than their univariate counterparts, and NWP RF is with the largest improvement. Generally, the edge of multivariate models is incrementally stronger with the raising forecast time. RF and LSTM, which perform unfavorably in univariate forecasts, excel in multivariate predictions. In particular, NWP LSTM dominates all cases, and its domination is reinforced over time. For four of the wind parks, the NWP models perform better in the term of MAE. The exception is Raggovidda wind park, for which the PE model has the lower MAE for all six-time steps than NWP MLP and CART. Meanwhile, the forecasting models produce the lowest MAE for Havoygavlen wind park.

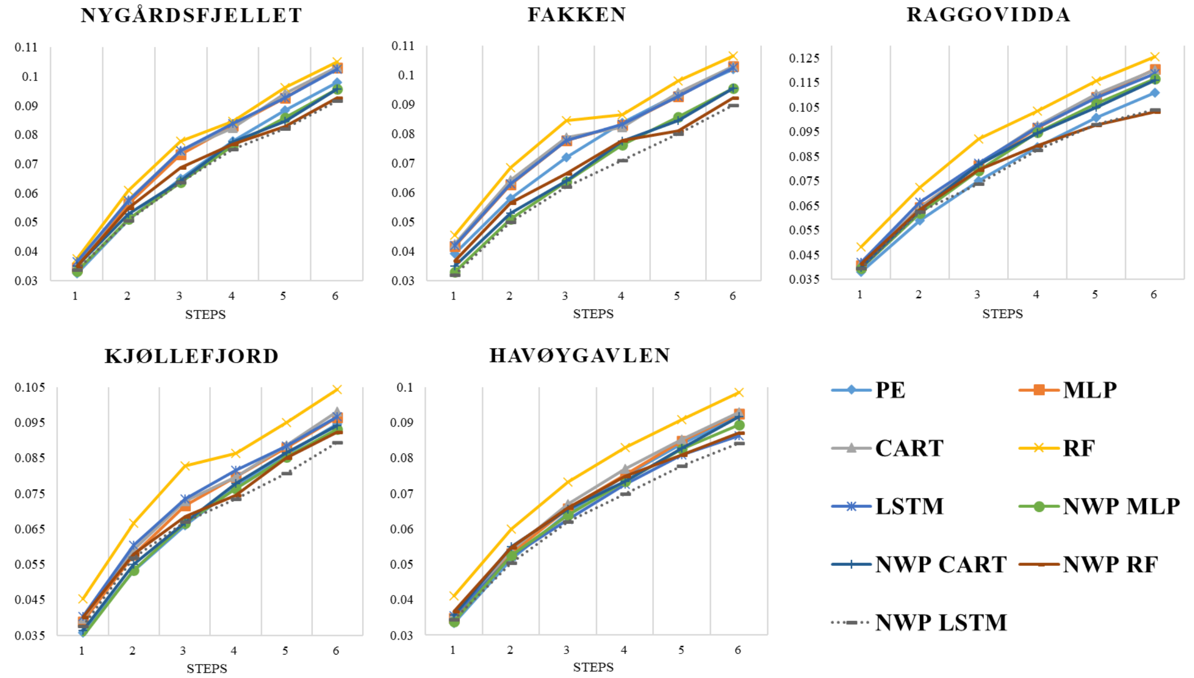


Fig. 4. The MAE comparisons of univariate and multivariate models for five wind farms

Numerical comparisons of average MAE for different models are shown in Table 3. The NWP LSTM model is ranked first for almost all wind parks. Adding the mesoscale NWP wind data can significantly increase forecasting algorithms' performance based on the same algorithm, especially for Fakken wind park, which reduces MAE by 16.14% and 16.54% concerning RF and LSTM, respectively.

Table 3. The comparisons of average MAE for different models

Wind park	PE v.s. NWP LSTM	MLP v.s NWP MLP	CART v.s NWP CART	RF v.s NWP RF	LSTM v.s NWP LSTM	NWP CART v.s NWP LSTM
Nygårdsfjellet	3.78%	8.41%	8.41%	11.02%	11.33%	3.18%
Fakken	14.16%	12.04%	11.94%	16.14%	16.54%	5.99%
Raggovidda	1.53%	2.71%	2.99%	14.89%	9.62%	6.96%
Kjøllefjord	1.81%	5.32%	4.82%	12.98%	8.11%	2.82%
Havøygavlen	5.46%	2.60%	1.92%	10.30%	3.02%	6.32%

The RMSE of the multivariate forecasts displays in Fig. 5. The trends of RMSE are similar to MAE's. Besides, growth rates of RMSE of all machine learning models, especially the NWP MLP, RF, and LSTM models, are lower than the PE model. Concerning each forecasting step, the multivariate models have lower RMSE than corresponding univariate models. For all the five wind parks, the multivariate LSTM performs best in terms of RMSE in nearly all the six predictive steps. Moreover, the Raggovidda wind park still has a higher RMSE compared to other wind parks. Meanwhile, the models still provide the lowest RMSE for Havøygavlen wind park.

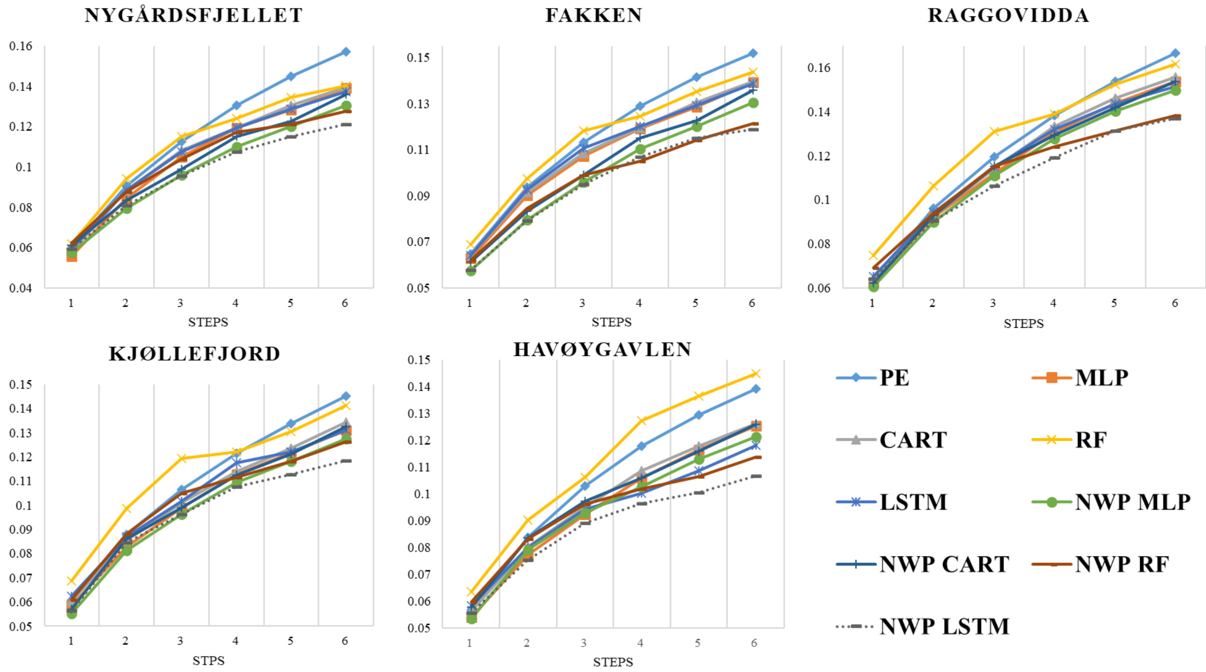


Fig. 5. The RMSE comparisons of univariate and multivariate models for five wind farms

Table 4 shows the comparisons of average RMSE and demonstrates the multivariate LSTM models are the best for all wind parks. It overperforms approximately an average of 14% better RMSE performance than the baseline PE model. The mesoscale NWP wind data provide positive information in the forecast algorithm.

Table 4. The comparisons of average RMSE for different models

Wind park	PE v.s. NWP LSTM	MLP v.s NWP MLP	CART v.s NWP CART	RF v.s NWP RF	LSTM v.s NWP LSTM	NWP CART v.s NWP LSTM
Nygårdsfjellet	16.69%	6.12%	4.27%	7.44%	9.56%	6.09%

Fakken	17.46%	8.25%	5.61%	14.83%	12.66%	7.21%
Raggovidda	12.14%	2.25%	1.57%	12.20%	7.53%	6.72%
Kjøllefjord	12.03%	2.95%	1.65%	10.33%	7.34%	5.33%
Havøygavlen	16.82%	1.56%	-0.36%	16.09%	6.43%	10.68%

6 CONCLUSION

This paper makes univariate and multivariate short-term wind energy forecasts for five wind parks inside the Norwegian Arctic region. Consequently, the following conclusions can be drawn.

For the univariate time series wind power prediction in these cases, the PE approach and machine learning methods do not have a considerable difference in performance. The SVR and MLP function equally well with the PE model. The machine learning algorithms that perform best in MAE are SVR and GA-SVR, whose average MAE is almost the same (0.18% and 0.10 % lower) as for the PE model. The machine learning algorithm that performs best in RMSE is MLP, whose average RMSE is 5.4% lower than the PE model. SVR, RBF, GA-SVR, and LSTM also have lower RMSE than the PE model has. This generally means the PE model has more large errors in the prediction procedure. Our result also validates the conclusion from research [19] in the wind engineering field. The conclusion is learning algorithms do not deliver on their promise for univariate time series prediction, and the classical statistical methods even perform better. The phenomena may be explained that for univariate series, the complex methods often overlearn the training set and create overfitting in the testing.

For the multivariate wind power forecasting in our cases, the model considers methodological or topographic factors by taking the mesoscale NWP wind data as inputs. Compared to the corresponding algorithm in the univariate case, the multivariate model has a lower MAE and results in a smaller RMSE. When the predictive time increases, the multivariate models are more stable than the PE model, especially in the metric of RMSE. These prove that the multivariate model entirely exceeds the PE model and the univariate model. Furthermore, the sophisticated ensemble and deep learning algorithm demonstrate their superiority in dealing with complex and multivariate pattern recognitions in complicated wind power forecasting problems.

The NWP wind data are generated with mesoscale (2.5km×2.5km), which is larger than the area of our wind parks. However, adding this local weather information can still obviously optimize the performance of forecasting models. The improvements of the penetration of NWP data in wind power prediction can be explained from two aspects: firstly, from the Bayesian theory [43], the introduced NWP wind information can provide a priori probability information to make more precise wind power (corresponding to posterior probability) predictions; secondly, NWP wind data can be regarded as the simulating wind conditions of the whole wind park, which add useful information in the predictive process.

In summary, based on our case studies, it is recommended to use statistical methods for short-term univariate wind power forecasting since learning algorithms involve parameter tuning and larger computational volumes without significantly better performance than statistical ones. It is advisable to include meteorological information, even if the weather data scale is relatively large, into the multivariate predictive models, where the advanced learning algorithms can be really effective with such autoregression combined with meteorological inputs.The replicability of these conclusions is established because the data from the five wind parks are relatively uncorrelated, and the machine learning regressions are built with sufficient considerations for the generalization of the models.

Moreover, we cannot find that each wind park's forecasting results significantly correlate with the site ruggedness. There are two possible reasons for this. First, the decisive, independent variable in the data-driven wind power prediction model is the prior value of power. According to Table 1, the standard deviation of the power time series does not become larger with increasing RIX. Second, because of the relatively large scale of the NWP model and the fact that the concerned wind farms are located near the sea, the effect of complex terrain is mitigated in the NWP grid. Therefore, to further investigate topography influence, the hydrodynamic modeling between topography and wind turbines in wind farms is needed. More interestingly, we find that the performances have some correlations with each wind park's capacity factor; the higher the capacity factor is, the lower the performance. i.e., the most inferior performance (Raggovidda) with the highest capacity factor of 48.40% and the best performance (Havøygavlen) with the lowest capacity factor of 25.46%. This implies that turbulence within the wind farm in complex terrain conditions may be the dominant factor, masking the effects of external large-scale weather.The phenomenon needs further investigations.

REFERENCES

1. Tollefson J. IPCC says limiting global warming to 1.5 C will require drastic action. *Nature* 2018; 562: 172-173.
2. Dudley B. BP Statistical Review of World Energy Statistical Review of World. *bp com* 2019.
3. Jung J and Broadwater RP. Current status and future advances for wind speed and power forecasting. *Renewable and Sustainable Energy Reviews* 2014; 31: 762-777.
4. Liu K, Zhang Y and Qin L. A novel combined forecasting model for short-term wind power based on ensemble empirical mode decomposition and optimal virtual prediction. *Journal of Renewable and Sustainable Energy* 2016; 8: 013104.
5. Birkelund Y, Alessandrini S, Byrkjedal Ø, et al. Wind power predictions in complex terrain using analog ensembles. 2018.

6. Samuel AL. Some studies in machine learning using the game of checkers. II—Recent progress. *IBM Journal of research and development* 1967; 11: 601-617.
7. Giebel G, Brownsword R, Kariniotakis G, et al. The state-of-the-art in short-term prediction of wind power: A literature overview. *ANEMOS plus* 2011.
8. Joyce T and Herrmann JM. A review of no free lunch theorems, and their implications for metaheuristic optimisation. *Nature-inspired algorithms and applied optimization*. Springer, 2018, pp.27-51.
9. Maldonado-Correa J, Solano J and Rojas-Moncayo M. Wind power forecasting: A systematic literature review. *Wind Engineering* 2019: 0309524X19891672.
10. Renani ET, Elias MFM and Rahim NA. Using data-driven approach for wind power prediction: A comparative study. *Energy Conversion and Management* 2016; 118: 193-203.
11. Godinho M and Castro R. Comparative performance of AI methods for wind power forecast in Portugal. *Wind Energy* 2020.
12. Wan C, Song Y, Xu Z, et al. Probabilistic wind power forecasting with hybrid artificial neural networks. *Electric Power Components and Systems* 2016; 44: 1656-1668.
13. Shao H, Deng X and Jiang Y. A novel deep learning approach for short-term wind power forecasting based on infinite feature selection and recurrent neural network. *Journal of Renewable and Sustainable Energy* 2018; 10: 043303.
14. Demolli H, Dokuz AS, Ecemis A, et al. Wind power forecasting based on daily wind speed data using machine learning algorithms. *Energy Conversion and Management* 2019; 198: 111823.
15. Maldonado-Correa J, Valdiviezo-Condolo M, Viñan-Ludeña MS, et al. Wind power forecasting for the Villonaco wind farm. *Wind Engineering* 2020: 0309524X20968817.
16. Yan J, Gao X, Liu Y, et al. Adaptabilities of three mainstream short-term wind power forecasting methods. *Journal of Renewable and Sustainable Energy* 2015; 7: 053101.
17. Torres J, Aguilar R and Zuñiga-Meneses K. Deep learning to predict the generation of a wind farm. *Journal of Renewable and Sustainable Energy* 2018; 10: 013305.
18. Cao Y, Hu Q, Shi H, et al. Prediction of wind power generation base on neural network in consideration of the fault time. *IEEJ Transactions on Electrical and Electronic Engineering* 2019; 14: 670-679.
19. Makridakis S, Spiliotis E and Assimakopoulos V. Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PloS one* 2018; 13: e0194889.
20. Bilal M, Birkelund Y and Homola M. High winds at nygårdsfjell. *Journal of Clean Energy Technologies,()* 2015.
21. Norway M. MetCoOp, https://www.met.no/en/projects/metcoop (accessed 25th, June 2020).
22. Collins SN, James RS, Ray P, et al. Grids in numerical weather and climate models. *Climate change and regional/local responses* 2013; 256.
23. Bremnes JB and Giebel G. Do regional weather models contribute to better wind power forecasts? : The Norwegian Meteorological Institute, 2017.
24. Roy SB. Simulating impacts of wind farms on local hydrometeorology. *Journal of Wind Engineering and Industrial Aerodynamics* 2011; 99: 491-498.
25. Ioffe S and Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. Cornell University Library. 2017.
26. Luo X. Size and power of tests for assessing weak stationarity of time series data: an empirical investigation. *schriftenreihe des studienGAnGs Geodäsie und GeoinformAtiK 2018–* 2018: 187.

27. Ghazali R, Hussain AJ, Nawi NM, et al. Non-stationary and stationary prediction of financial time series using dynamic ridge polynomial neural network. *Neurocomputing* 2009; 72: 2359-2367.
28. Li B, Zhang J, He Y, et al. Short-term load-forecasting method based on wavelet decomposition with second-order gray neural network model combined with ADF test. *IEEE Access* 2017; 5: 16324-16331.
29. Raschka S and Mirjalili V. *Python machine learning*. Packt Publishing Ltd, 2017.
30. Sagi O and Rokach L. Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2018; 8: e1249.
31. Smola A. Support vector machines, regularization, optimization, and beyond. MIT press, Cambridge, MA, USA, 2002.
32. Walters-Williams J and Li Y. Comparative study of distance functions for nearest neighbors. *Advanced Techniques in Computing Sciences and Software Engineering*. Springer, 2010, pp.79-84.
33. Panchal G, Ganatra A, Kosta Y, et al. Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers. *International Journal of Computer Theory and Engineering* 2011; 3: 332-337.
34. Breiman L, Friedman J, Stone CJ, et al. *Classification and regression trees*. CRC press, 1984.
35. Babar B, Luppino LT, Boström T, et al. Random forest regression for improved mapping of solar irradiance at high latitudes. *Solar Energy* 2020; 198: 81-92.
36. Friedman JH. Stochastic gradient boosting. *Computational statistics & data analysis* 2002; 38: 367-378.
37. Whitley D. A genetic algorithm tutorial. *Statistics and computing* 1994; 4: 65-85.
38. Hochreiter S and Schmidhuber J. Long short-term memory. *Neural computation* 1997; 9: 1735-1780.
39. Kohavi R. Proceedings of the 14th international joint conference on Artificial intelligence. 1995.
40. Chikkerur S, Sundaram V, Reisslein M, et al. Objective video quality assessment methods: A classification, review, and performance comparison. *IEEE transactions on broadcasting* 2011; 57: 165-182.42
41. Lahouar A and Slama JBH. Hour-ahead wind power forecast based on random forests. *Renewable energy* 2017; 109: 529-541.
42. Xiaoyun Q, Xiaoning K, Chao Z, et al. Short-term prediction of wind power based on deep long short-term memory. In: *2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)* 2016, pp.1148-1152. IEEE.
43. Bernardo JM and Smith AF. *Bayesian theory*. John Wiley & Sons, 2009.

Paper III. B.

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

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

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 (Postprint)

Hao Chen, Yngve Birkelund, Fuqing Yuan

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

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 The 4th International Conference on Electrical Engineering and Green Energy.

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].

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 95m. 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:

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

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 equation (3).

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

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

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_o: The means of these populations are equivalent; H_a: At least one does not equal the other. Their test statistics are as below:

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

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

The Tukey method for CIs is expressed as:

$$(\bar{Y}_1 - \bar{Y}_{2\cdot}) \pm \frac{q_{k,n-k,1-\alpha}}{\sqrt{2}} \cdot \sqrt{MSE} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \tag{6}$$

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,

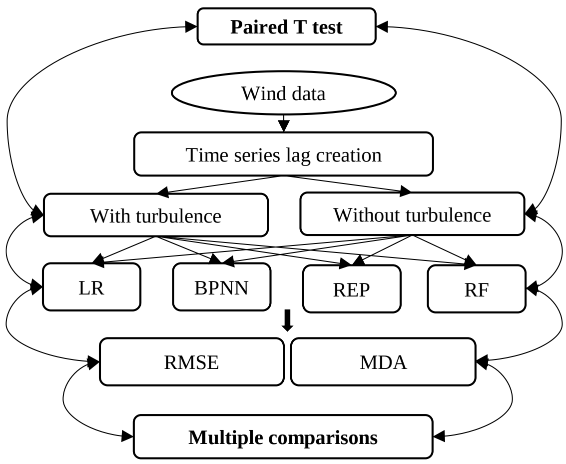


Fig. 1. Procedure for wind forecasts and statistical tests.

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.

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. Tables 1 shows RMSEand 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 Tables 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 hours (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 Tables 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.

References

[1] Tian Z, Ren Y and Wang G. (2018) "Short-term wind power prediction based on empirical mode decomposition and improved extreme learning machine." *Journal of Electrical Engineering & Technology* 13 1841-51.

[2] Maldonado-Correa J, Solano J and Rojas-Moncayo M. (2019) "Wind power forecasting: A systematic literature review." *Wind Engineering* 0309524X19891672.

[3] Singh S N and Erlich I. (2008) "Strategies for wind power trading in competitive electricity markets." *IEEE transactions on energy conversion* 23 249-56.

[4] Hong D, Ji T, Li M and Wu Q. (2019) "Ultra-short-term forecast of wind speed and wind power based on morphological high frequency filter and double similarity search algorithm." *International Journal of Electrical Power & Energy Systems* 104 868-79.

[5] Gangui Y, Yu L, Gang M, Yang C, Junhui L, Jigang L and Lei M. (2012) "The ultra-short-term prediction of wind power based on chaotic time series." *Energy Procedia* 17 1490-6.

[6] Zhang Z Z, Zou J X and Zheng G. (2011) "Ultra-short term wind power prediction model based on modified grey model method for power control in wind farm." *Wind Engineering* 35 55-67.

[7] Shi K, Qiao Y, Zhao W, Wang Q, Liu M and Lu Z. (2018) "An improved random forest model of short-term wind-power forecasting to enhance accuracy, efficiency, and robustness." *Wind energy* 21 1383-94.

[8] Lee J, Wang W, Harrou F and Sun Y. (2020) "Wind Power Prediction Using Ensemble Learning-Based Models." *IEEE Access* 8 61517-27.

[9] Nielson J, Bhaganagar K, Meka R and Alaeddini A (2020) "Using atmospheric inputs for Artificial Neural Networks to improve wind turbine power prediction." *Energy* 190 116273.

[10] Li F, Ren G and Lee J. (2019) "Multistep wind speed prediction based on turbulence intensity and hybrid deep neural networks." *Energy Conversion and Management* 186 306-22.

[11] Jacobson M Z and Archer C L. (2012) "Saturation wind power potential and its implications for wind energy." *Proceedings of the National Academy of Sciences* 109 15679-84.

[12] Kaiser K, Langreder W, Hohlen H and Højstrup J. (2007) *Wind Energy*, Springer.

[13] Lubitz W D. (2014) "Impact of ambient turbulence on performance of a small wind turbine." *Renewable Energy* 61 69-73.

[14] Türk M and Emeis S. (2010) "The dependence of offshore turbulence intensity on wind speed." *Journal of Wind Engineering and Industrial Aerodynamics* 98 466-71.

[15] Trenkler G. (1996) "Methods of multivariate analysis". Wiley series in probability and mathematical statistics: Probability and mathematical statistics section." *Computational Statistics & Data Analysis* 22 334-5.

[16] Samadianfard S, Hashemi S, Kargar K, Izadyar M, Mostafaeipour A, Mosavi A, Nabipour N and Shamshirband S. (2020) "Wind speed prediction using a hybrid model of the multi-layer perceptron and whale optimization algorithm." *Energy Reports* 6 1147-59.

[17] Zhao Y and Zhang Y. (2008) "Comparison of decision tree methods for finding active objects." *Advances in Space Research* 41 1955-9.

[18] Ho T K. (1995) "Random decision forests. " *IEEE Proceedings of 3rd international conference on document analysis and recognition*.

[19] Babar B, Luppino L T, Boström T and Anfinsen S N. (2020) "Random forest regression for improved mapping of solar irradiance at high latitudes." *Solar Energy* 198 81-92.

[20] Kleinbaum D G, Kupper L L, Nizam A and Rosenberg E S. (2013) *Applied regression analysis and other multivariable methods*, Nelson Education.

Paper IV.

Data-augmented sequential deep learning for wind power forecasting

https://doi.org/10.1016/j.enconman.2021.114790

Data-augmented Sequential Deep Learning for Wind Power Forecasting (Postprint)

Hao Chen[1*], Yngve Birkelund[2], Qixia Zhang[3]

[1]*Department of Technology and Safety, UiT The Arctic University of Norway, Tromsø 9019, Norway*

[2]*Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø 9019, Norway*

[3]*Huazhong University of Science and Technology, Wuhan 430074, China*

Abstract

Accurate wind power forecasts play a critical role in the operation of wind parks and the dispatch of wind energy into the power grid. With excellent automatic pattern recognition and nonlinear mapping ability for big data, deep learning is increasingly employed in wind power forecasting. However, salient realities are that in-situ measured wind data are relatively expensive and inaccessible and correlation between steps is omitted in most multistep wind power forecasts. This paper is the first time that data augmentation is applied to wind power forecasting by systematically summarizing and proposing both physics-oriented and data-oriented time-series wind data augmentation approaches to considerably enlarge primary datasets, and develops deep encoder-decoder long short-term memory networks that enable sequential input and sequential output for wind power forecasting. The proposed augmentation techniques and forecasting algorithm are deployed on five turbines with diverse topographies in an Arctic wind park, and the outcomes are evaluated against benchmark models and different augmentations. The main findings reveal that on one side, the average improvement in RMSE of the proposed forecasting model over the benchmarks is 33.89%, 10.60%, 7.12%, and 4.27% before data augmentations, and increases to 40.63%, 17.67%, 11.74%, and 7.06%, respectively, after augmentations. The other side unveils that the effect of data augmentations on prediction is intricately varying, but for the proposed model with and without augmentations, all augmentation approaches boost the model outperformance from 7.87% to 13.36% in RMSE, 5.24% to 8.97% in MAE, and similarly over 12% in QR90. Finally, data-oriented augmentations, in general, is slightly better than physics-driven ones.

Keywords: Renewable energy; Wind power forecasting; Data augmentation; Deep learning; Encoder-decoder networks; Big data

Key abbreviations

$\hat{P}_{i+n}$	*n* timestep ahead predicted wind power
P_i	Measured wind power
v_i	Measured wind speed
u_{i+n}	*n* timestep ahead wind speed calculated from weather model
m	Sample number of the testing set
Cap	Designed capacity of the wind turbine
T	Statistic of paired T-test
F	Statistic of paired Friedman test
BA	Bionic optimized neural networks constructed Adaboost
DA#	Physics-oriented data augmentation strategy - number #
ED	Encoder-Decoder
EDLSTM	Proposed Encoder-Decoder Long Short-Term Memory neural networks
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MSE	Mean Square Error
NLP	Natural Language Processing
NN	Three-layer backpropagation Neural Networks
NWP	Numerical Weather Prediction
PA#	Data-oriented data augmentation strategy – strategy number #
PI	Prediction Interval
PR	Persistence model
QR90	Qualification Rate at the 90% threshold
RMSE	Root Mean Square Error
RNN	Recurrent Neural Networks
seq2seq	Sequence-to-Sequence
STD	Standard deviation
T#	Wind turbines with different terrain – turbine number #

1. Introduction

Wind is a renewable, sustainable, and environmentally friendly energy resource. As wind technology has developed in recent years, wind energy has received attention from a growing number of countries for its low-cost operation and maintenance, small turbine footprint, flexibility in development scale, and rapidly decreasing electricity generation costs. [1]

Meanwhile, massive electricity generated by wind energy is volatile, intermittent, and with low power density. These features influence the power production of generation companies, the balance of the grid and may profoundly jeopardize its security [2]. In a large-scale grid-connected system involving wind power, an unplanned load increase or an unscheduled wind power decrease will cause a supply-demand imbalance when thermal power or hydropower ceases generation or is insufficient. [3] Hence, the uncertainty in wind power production enlarges the required reserve capacity of the system. An accurate wind power forecast minimizes the spare capacity and enables optimal dispatch of power in systems with wind power generation. Furthermore, an effective prediction serves as a basis for wind parks to engage in generation bidding, determines a reasonable charging and discharging strategy for energy storage, and lowers the occurrence and duration of wind curtailments.

Wind power forecasting methodology is generally divided into physical, statistical, and hybrid approaches. [4] The first predicts wind power by extensive numerical computation of physical equations. It is based on fluid dynamics and uses Numerical Weather Prediction (NWP) data such as wind speed and pressure, and geoinformation like ground roughness and altitude. The method performs best in medium or long-term forecasting and applies to the wind resource assessment of new wind parks that lack historical observations. The statistical approach aims to establish linear or nonlinear patterns within wind data sequences that can be utilized in forecasting. In particular, machine learning-based wind power forecasting methods developed in recent years are widely applied. The hybrid approach is a combination of the former categories and has shown its edge profoundly. [5]

In 2006, Hinton et al. successfully trained deep neural networks (i.e., artificial neural networks with several hidden layers) and achieved excellent performance on multiple datasets, [6] which signified the birth of deep learning. Since then, deep learning techniques based on neural networks of different designs have flourished and solved long-standing challenges, such as voice and image recognition and generation, preliminary implementation of autonomous driving, etc. [7]. Recently, the application of deep learning to energy science has also become popular because of its powerful auto-pattern recognition and nonlinear mapping capabilities. [8] The two major drivers of deep learning evolution are progressive computational capabilities and the influx of big data. It is generally agreed that larger datasets yield better deep learning models. [9]

The effectiveness of deep supervised learning relies on the volume and quality of labeled training data as well as the topology and parameters tuning of deep networks. [10] Notably, an effective solution to establish large sets of training data is data augmentation, since the training set typically lacks a sufficient number of manually labeled samples. Especially in wind energy, it is generally challenging to acquire high-quality and long-duration meteorological and power production data.

Data augmentation is a technique to make supervised machine learning more efficient. It extends the amount of available training data by adding modified versions of existing data or new data generated based on existing data. The effect is to regularize the deep learning model and assist in mitigating overfitting during deep training, thereby improving the generalizability and ubiquity of the learned models. Overfitting is a phenomenon that occurs as a learner learns a function with extraordinarily large variance, such as perfectly fitting the training data. Generalizability defines the difference in performance when a model is assessed in relation to data in the training set previously seen compared to previously unseen data in the testing set. [11]

Essentially, using multi-inputs to make multistep wind power forecasting can be regarded as a Sequence-to-Sequence (seq2seq) prediction that is framed as a mapping of multiple inputs to multiple time-series outputs. It was demonstrated that the seq2seq model "*approaches or surpasses all currently published results*" [12] in Natural Language Processing (NLP), like Google Translate, and recently it has also shown its promise in renewable energy forecasting. [13] , [14] The Encoder-Decoder (ED) Recurrent Neural Networks (RNN) has successfully handled seq2seq problems [15] and exhibits state-of-the-art performance in the area of text translation that is fundamentally a time-series problem.

1.1. Previous work review

In computer science research, there are several developed methodologies in data augmentation. [16] Shorten and Khoshgoftaar [11] systematically presented current imagery data augmentation methods, their promising advances, and methodologies used to implement them to boost the performance of imagining deep learning tasks. Cubuk et al. [17] investigated several commonly used image recognition datasets and designed an augmentation strategy that learns from the datasets. The strategy consists of many sub-strategies and is automatically selected in the model training process and helps gain 0.4% to 0.6% imagine classification accuracy on different datasets. But the data augmentation technique is mainly in the field of image recognition and has received little attention in sequence analysis. DeVries and Taylor [18] summarized and utilized interpolation and extrapolation, etc., and domain-agnostic approach to reach the predictions with deep learning for time-series datasets, and tentatively proved the techniques are timely and effective in some supervised learning problems.

Deep learning techniques have got much attention from researchers in renewable energy forecasting.[8] With its distinctive automatic nonlinear recognition capabilities, deep learning has gradually emerged as an important approach to the challenge of forecasting sharply volatile wind power. [5], [19] Yildiz et al. [20] extracted wind datasets with features with variational mode decomposition and converted these features into images. Then the images were handled by an improved residual-based deep convolutional neural network to forecast wind power for a wind park in Turkey. The edge of the proposed process was proved by a comparison between some existing well-used large networks. Kisvari et al. [21] constructed a framework consisting of data preprocessing, anomaly detection, feature engineering, and gated recurrent deep learning models for wind power prediction and demonstrated that the framework offered more effective predictions than traditional recurrent neural networks. Shahid et al. [22] piled up Long Short-Term Memory (LSTM) units into a large network and tunes the network by using the genetic algorithm to forecast wind power validated the statistical advantage of the network over a single unit by the Wilcoxon Signed-Rank test. Memarzadeh et al.[23] applied a bionic algorithm, wavelet transform, feature selection, and LSTM networks to forecast wind power of two wind parks in Spain and Iran, and showed the effectiveness of the proposed method by comparison with benchmark neural networks.

While numerous wind power models based on a hybrid of traditional data methodologies and deep learning have been developed and advanced in forecasting for many sites, nevertheless, further sophistication of forecasting models may render the results specific, i.e., wind power forecasts are restricted to a certain category of terrain and weather features and difficult to be generalized and not be easily employed because their consisting techniques such as signal processing, feature engineering, etc. require a prolonged and special training to master. Lipu et al. [24] also summarized the most recent progress of wind power forecasting using artificial intelligence and pointed out the issues and challenges in the field. The challenges include many various data preprocessing techniques for diverse wind data, model structure, and optimization, etc. In particular, Reichstein et al. [25] recommended that more attention should be given to Earth system science problems to the coupled data approaches with physical phenomena and deep learning methods themselves, rather than building more complex traditional methods-based models.

In the present study, in the contrast, we return to the physical process of wind power generation, the statistical characteristics of wind data, and the nature of deep learning to approach the forecasting problem. After synthesizing numerous data augmentation methodologies and drawing on multiple state-of-the-art advances in sequential data prediction, the robust and efficacious encoder-decoder deep neural networks with stacking LSTM units are proposed for wind turbine power forecasting in the Arctic.

1.2. Contributions

Leveraging the aforementioned literature review, attention is paid to a wind park, inside the Arctic, in complex terrain. The principal contributions of the present study paper are as follows:

1. This paper systematically applies data augmentation to wind power forecasting for the first time. Specifically, eight time-series data augmentation approaches are proposed according

to physical characteristics of wind energy and statistical properties of data in wind engineering. The approaches are implemented in four benchmarking models and proposed advanced deep learning models. The methodology is particularly suitable for new wind parks that have a short period of operation and therefore a limited amount of accumulated data. It enables to fully and automatically deepen the information and value of these limited data.

2. We exhaustively develop a seq2seq deep learning predictive end-to-end model with inputs of historical wind speed and power data and wind speed from NWP as well as simultaneously interrelated outputs of multistep, futuristic wind power. The model is based on an encoder-decoder constructed with LSTM and shows its superiority in forecasting power.
3. It is demonstrated that the impact of various augmentation approaches is different in each forecasting algorithm. Augmentations somewhat increase linear, like persistence, model errors. Nonetheless, augmentations improve the performance, most notably the proposed deep learning model, of neural networks-based algorithms, where data-oriented augmentations generally contribute greater than physics-oriented ones.
4. The data augmentations combined with the proposed and benchmark forecasting models are utilized to predict power generated by five turbines in various landscapes. The results are analyzed by rigorous statistical methods and indicate that the augmentations and the proposed forecasting model have wind engineering values and potentially extensive applicability in other energy fields.

The architecture avenue opens the article with an introduction on wind energy forecasting and its deep learning utilization status quo as well as contributions presented in Section 1. Section 2 illustrates the principle of wind power generation and the utilized data and scheme. Section 3 delves into proposed data augmentation techniques and a novel predictive deep neural network. Section 4 provides detailed experiment procedures and model assessment metrics. In Section 5, hierarchical experimental results and discussions, from comparisons of models themselves to data augmentation approaches, are presented. Finally, the main findings, research outlooks, and derivative policy recommendations are demonstrated in Section 6.

2. Data preparation and forecast scheme

Wind power generation is a conversion from wind energy to electricity. Ideally, the output generation of a wind turbine is expressed as in (1):

$$P = \begin{cases} 0 & v < v_{min} \\ P_v(C_P, \rho, A; v) & v_{min} < v < v_n \\ P_r & v_n < v < v_{max} \\ 0 & v > v_{max} \end{cases} \tag{1}$$

where P is the output power of the wind turbine (W); P_v (.), typically proportional to the cubic of the wind speed, is the wind curve function at the speed interval, C_P means wind energy utilization efficiency; ρ is the air density (kg/m^2); A is the effective area swept by turbine blades (m^2), v denotes the wind speed (m/s); $v_{\min}$, $v_{\max}$, and v_{n} respectively are cut-in, cut-off and rated wind speed. P_r is turbine rated wind power. From (1), the output of a wind turbine is mainly influenced by the third power of wind speed, air density, and swept area.

The study centers on the wind turbine, 3.0 MW Vestas V90, electricity production of a wind park, named Fakken, with an installed capacity of 54 MW with 18 turbines, average annual production is 139 GWh in the Arctic region.

Fig. 1. Fakken wind park located in northern Norway

Wind is predominantly influenced by the terrain; wind anomalies occur when wind moves through these areas. The influence is dependent on the height and width of the barriers. The terrain of Fakken wind park is with low and flat hills and narrow valleys, and towards a fjord. Fig. 1 is a photograph of the wind park in operation at a close distance taken by the authors in May 2021 at sea, and the nearby mountain reaching around 900 meters above sea level.

The timescale of data in this study is from 0:00 1st January 2017 to 23:50 31st December 2017. Raw wind speed and power data of each turbine, 10 mins temporal resolution and recorded by Supervisory Control And Data Acquisition SCADA, are supplied by a local wind energy operator. The NWP wind speed data, calculated by the Meteorological cooperation on operational Ensemble Prediction System (MEPS) NWP model, are with 2.5 km horizontal resolution that is taken as the mesoscale. The model, operating by the Norwegian Meteorological Institute, updates at 00, 06, 12, and 18 UTC, and its forecasts for the next 66 hours are available around 1 h 15 minutes later. The wind speed data sequences from NWP comprise the nearest accessible weather prediction data.

To verify the generality and portability of the proposed methodology, five wind turbines separately situated in different topographic conditions in the wind park are selected as study subjects. Moreover, wind measurements are taken at the turbine nacelle, which is 80 meters about the ground. Their topographic features and statistics of annual in-situ measured wind speed and power are shown in Table 1.

Table 1. The terrain and statistics of wind turbines.

Wind Turbine	Terrain	Wind power				Wind speed			
		Mean[kW]	STD[kW]	Skew	Kur	Mean[m/s]	STD[m/s]	Skew	Kur
T1	Plateau	825.58	990.43	1.01	-0.35	3.98	5.15	1.18	1.31
T2	Valley	826.19	987.92	0.95	-0.43	3.91	5.06	1.10	1.12
T3	Lakeside	738.37	914.33	1.23	0.28	3.55	4.65	1.33	2.01
T4	Hilltop	804.40	971.86	1.04	-0.25	4.02	5.27	1.27	1.71
T5	Seaside	783.86	950.42	1.06	-0.14	3.93	5.10	1.19	1.49

Note: STD is standard deviation, Skew is skewness and Kur is relative kurtosis (actual kurtosis minus 3).

Statistically, wind power forecasting can be regarded as a multivariable regression problem, in which wind power time series are autoregressed, and wind speed serves as the supplementing information to the autoregression. Updating the wind speed from NWP of the predicted time, the current information, is also the key feature in the prediction since according to an extensively cited reference by Giebel and Kariniotakis [26], forecasting wind power beyond three to six hours typically requires consideration of information on NWP wind speed at the moment of prediction. In this study, we chose measured data of the previous six hours to make multistep forecasts for the wind power from the next six to twelve hours with the assistance of wind speed from NWP.

The fundamental multistep forecasting model $f(.)$ with timestep $i+n$ is described as:

$$\hat{P}_{i+n} = f\left(P_{i-j}; v_{i-j}; u_{i+n}\right) + \varepsilon_n \tag{2}$$

where i represents the base current time i=1, 2, ..., 7, and with each i, j=0, 1, ..., 6. $\hat{P}_{i+n}$ is n timestep ahead predicted wind power, $n \in \{6,7,8,9,10,11,12\}$, v is the wind speed observed in the turbine, u represents the wind speed calculated from the mesoscale NWP wind model for the site. ε_n is the error of the forecasting model.

Since the scopes of wind power and speed are not the same, it is beneficial to rescale the raw data into a new set with a similar scale. Data rescaling techniques can accelerate convergence speed and improve algorithms' accuracy of neural networks. [27]

3. Methodology

3.1. Wind data augmentation

In practice, testing errors need to be continuously reduced along with training errors to construct meaningful deep learning models. Data augmentation is a phenomenally robust approach to accomplish this aim. It embarks on overfitting from the origin, the training data themselves, of the problem, assuming that further information can be retrieved from the source dataset.

Based on know-how in wind energy technology and state-of-the-art data science, we divide the techniques for augmenting wind data for forecasting with robust and efficient deep learning into two categories: physics-oriented and data-oriented.

3.1.1. Physics-oriented approaches

Inspired by the physics of wind power engineering, we propose three strategies to augment training set data for forecasting models. The first is the explicit perturbation of the wind power curve according to Eq. (1). The second is the implicit perturbation based on the difference between the numerical weather predicted wind speed of the wind park area and the actual measured wind speed of turbines. The third considers the operational data of the other wind turbines in the vicinity of the studied wind turbines. These three physics-oriented approaches are shortened as PA1, PA2, and PA3, respectively.

PA1: Considering the wind speed as the independent variable and differentiating Eq. (1), the following Eq. (3) is obtained.

$$dP = \begin{cases} 0 & v < v_{min} \\ P_v'(C_P, \rho, A; v)dv & v_{min} < v < v_n \\ 0 & v_n < v < v_{max} \\ 0 & v > v_{max} \end{cases} \tag{3}$$

from Eq. (1), it is observed that when v is in the cut-in and rated wind speed interval, the derivative of the power curve, the ratio of tiny variations in wind turbine power and wind speed, is proportional to the quadratic of this point wind speed. Therefore, according to Eq. (3), it is possible to artificially adhere a slight random perturbation in a wind speed point in the interval and calculate the corresponding power variation in accordance with the speed.

PA2: According to Eq. (2), the input to the power forecasting model contains the wind speed from measurements and the NWP model, but they correspond to different time stamps when entering the model. Since NWP datasets also have wind speeds that correspond to the same time stamps as the measured wind speeds, and there is no significant difference in wind speed probability distribution from two wind speed resources in the wind park based on our previous study [28]. So, we resort to a random replacement strategy with a fixed probability to replace the wind speeds in the measured datasets with the correspondent NWP wind speeds.

PA3: Since the neighboring turbines to the target turbine have similar wind conditions in operation. Therefore, adopting the measured wind speed of the neighboring turbine with a specific probability to replace the target turbine could be a strategy to augment the target wind speed dataset.

3.1.2. Data-oriented approaches

The proposed taxonomy for the data-oriented methods for wind power forecasting is enlightened by the feature space expansion, signal processing, and machine learning techniques. It consists of five approaches. DA1: Various simple interpolation and extrapolation methods are used to obtain data on larger time scales. DA2: Implements noise to the original dataset. DA3: Sequential augmentation approaches, named geometric transformations, draw on image processing, symmetry or flipping, translation, and random erasing. DA4: Methodology used for decomposition in time-series data. DA5. Scenario generation methods for the single turbine include statistical and machine learning generation.

DA1: Averaging is usually required to calculate the data in hourly units as the original measured dataset is in ten-minute increments. The new hourly data can be acquired by performing some interpolation or extrapolation modification to this averaging process. The new averaging is defined as:

$$\boldsymbol{x}_{\boldsymbol{t}}' = \sum_{j=1}^{6} \omega_j \boldsymbol{x}_j \tag{4}$$

where $\boldsymbol{x}_{\boldsymbol{t}}'$ is the hourly data and $\boldsymbol{x}_j$ donates the raw 10-mins data. ω_j is the stochastic weight that fulfills: $\sum_{j=1}^{6} \omega_j = 6, (-0.3 \leq \omega_j \leq 1.3)$, which when $\omega_j < 0$ is extrapolation while $\omega_j \geq 0$ means interpolation.

DA2: Another simple, probably the simplest, method of data augmentation is the addition of white noise, following the standard normal distribution, to data. A wind power forecasting study considered noise in data as a detrimental factor for prediction and removed it by signal processing. [29] Nonetheless, in machine learning research, applying noise to the neural network's inputs increases the generalizability of the networks. [16] The noise injection is determined with a scaling parameter $\boldsymbol{\delta}$:

$$\boldsymbol{x}_{\boldsymbol{t}}' = \boldsymbol{x}_{\boldsymbol{t}} + \boldsymbol{\delta X}, \boldsymbol{X} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\sigma}_{\boldsymbol{i}}) \tag{5}$$

where $\boldsymbol{x}_{\boldsymbol{t}}'$ is the enhanced data and $\boldsymbol{x}_{\boldsymbol{t}}$ donates the original hourly data.

DA3: Geometric transformations are among the initial data augmentation methods with excellent effectiveness in deep learning for image recognition, such as flipping, cropping, and color transformations. [11] Based on the characteristics of the measured wind speed time series and referring to image geometric augmentations, we stochastically opt for, 10% respectively, symmetry along with the average point, substitution of prior or posterior values, and stochastic erasing of some data.

DA4: Wind power forecasting is known mathematically as a special time series problem. Ordinarily, the time series x_t can be decomposed into base α_t, trend τ_t, season s_t, and residual γ_t parts as in Eq. (6).

$$x_t = \alpha_t + \tau_t + s_t + \gamma_t \text{ , } t = 1,2,\ldots, N \tag{6}$$

The extensively implemented approach is firstly based on the time-domain figure of the time series or its Fourier analysis to obtain its period corresponding to seasonality, and then decomposes the time series with the loess smoothing technique [30], a locally weighted autoregression, into the above four components. The weights of these four components are subsequently and stochastically adjusted by Eq. (7) to form an augmented series.

$$x_t' = \omega_1\alpha_t + \omega_2\tau_t + \omega_3 s_t + \omega_4\gamma_t \text{ , } \sum_{i=1}^{4}\omega_i = 4,\ 0.9 \le \omega_i \le 1.1 \tag{7}$$

DA5: The data augmentation methodologies described above all involve randomness, data selections, and/or weight adjustments, so they are relatively independent of the data and require considerable manual fine-tuning. Wind power scenario generation is an effective tool to resolve uncertainties in stochastic planning of the energy system with the integration of wind power. [31] Classical and advanced statistical methods and machine learning models are broadly employed [32] to predict wind power scenarios. Intrinsically, these models profile conditional distributions of time series by assuming that the current value depends on previous points: a new time series may be generated from the learned conditional distributions provided that original series values are perturbed in some way.

3.2. Encoder-decoder LSTM deep networks

RNN has achieved tremendous success and wide application in numerous sequence applications. [16] RNN is designed to process learning tasks with sequential data. 'Recurrent' means the current output is related to the previous output. The nodes in hidden are structurally connected to each other to reach inputs of the hidden layers includes not only outputs of the input layer but also ones of the previous-time hidden layers.

Among the RNN network structures, the most extensively used and highly successful model is the LSTM network, with a kind of unique memory unit in its hidden layers and is generally more expressive of long-short time dependencies than the other RNNs. [33] Typically, the LSTM unit consists of three gates, i.e., input gate, forget gate, and output gate. There are three primary internal phases of the unit. The first is forget phase, which retains the important information coming in from the previous node and forgets the unimportant details. The next phase is the selective memory phase, which optionally remembers inputs of this phase. Finally, an output phase determines which ones should be treated as outputs of the current state. Mathematically, the long-short memory unit can be expressed as [34]:

$$
\begin{aligned}
&\boldsymbol{i}_t := \sigma\ (\boldsymbol{W}_{xi}\boldsymbol{x}_t + \boldsymbol{W}_{hi}\boldsymbol{h}_{t-1} + \boldsymbol{b}_i)\,,\\
&\boldsymbol{f}_t := \sigma\ \left(\boldsymbol{W}_{xf}\boldsymbol{x}_t + \boldsymbol{W}_{hf}\boldsymbol{h}_{t-1} + \boldsymbol{b}_f\right),\\
&\boldsymbol{o}_t := \sigma\ (\boldsymbol{W}_{xo}\boldsymbol{x}_t + \boldsymbol{W}_{ho}\boldsymbol{h}_{t-1} + \boldsymbol{b}_o),\\
&\tilde{\boldsymbol{c}}_t := \tanh\ (\boldsymbol{W}_{xc}\boldsymbol{x}_t + \boldsymbol{W}_{hc}\boldsymbol{h}_{t-1} + \boldsymbol{b}_c),\\
&\boldsymbol{c}_t := \boldsymbol{f}_t \odot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \odot \tilde{\boldsymbol{c}}_t,\\
&\boldsymbol{h}_t := \boldsymbol{o}_t \odot \tanh(\boldsymbol{c}_t)\,.
\end{aligned}
\tag{8}
$$

where $\boldsymbol{x}_t$ is the input and $\boldsymbol{h}_{t-1}$ is the hidden state of the previous timestep. $\boldsymbol{i}_t$, $\boldsymbol{f}_t$, and $\boldsymbol{o}_t$ are input, forget, and output gates, $\boldsymbol{W}_{..}$ denotes the corresponding weight parameter, and $\boldsymbol{b}_.$ is the corresponding bias parameter. $\tilde{\boldsymbol{c}}_t$ is the candidate memory cell, $\boldsymbol{c}_t$ is the memory cell, and $\boldsymbol{c}_{t-1}$ is its previous time step state. $\boldsymbol{h}_t$ is the hidden state. σ (.) is the sigmoid function, tanh (.) is hyperbolic tangent function, and $\odot$ represents the pointwise multiplication.

The encoder-decoder LSTM is a type of EDRNN network designed to deal with seq2seq, and its architecture is innovative in terms of sequence embedding, i.e., the usage of a reading-in and exporting-out fixed-size sequences. The encoder-decoder LSTM includes an input layer, LSTM based encoder and decoder, and an output layer in this study. The LSTM unit achieves the extraction and utilization of important information in the sequence through its gate controls. The encoder reads input sequences and encodes them into fixed-length vectors by the weight of each time step with a context vector. The decoder decodes these fixed-length vectors and outputs predicted sequences. The fixed-length context vector introduces a mechanism called Attention, which enables highly summarize and highlight the information learned by the encoder and uses it as input to the decoder for translation. The encoder and decoder networks are mutually independent, which indicates that their LSTM units do not share parameters during the process of networks training.

3.3. Proposed deep EDLSTM for wind power forecasting

According to Eq. (2), wind power prediction involves autoregression, multiple sources of wind speed, and nonlinear functional relationships, all of which may lead to the application of EDLSTM networks. In addition, multistep wind power forecasting is appropriate to be handled as a seq2seq problem since the historical data of the inputs are linked and interactive. Therefore, a deep, stacked multilayers EDLSTM, shorten as EDLSTM, is proposed and utilized to extract the implicit features from layer to layer. The detailed deep EDLSTM employed in this article is illustrated in Fig. 2.

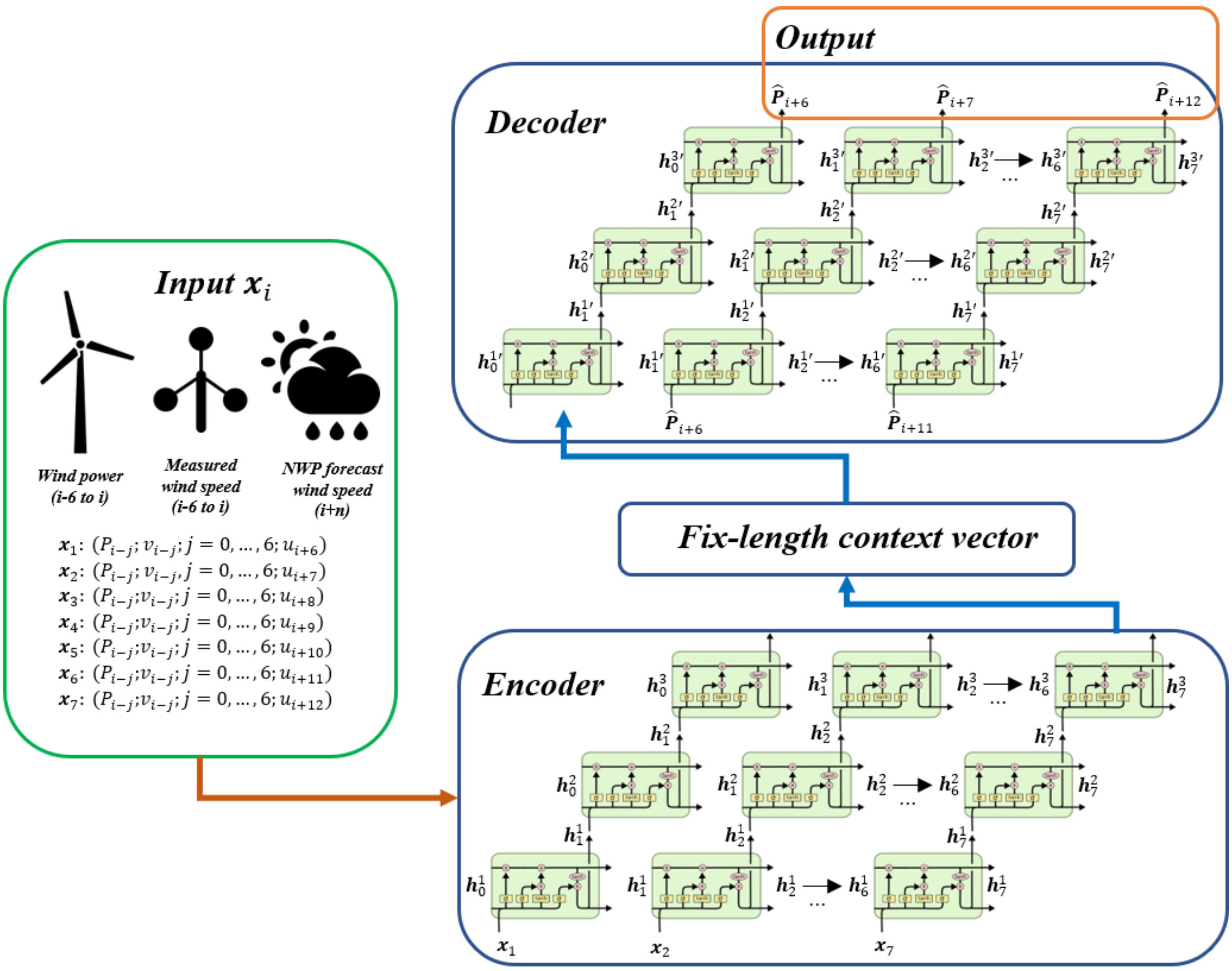


Fig. 2. The structural diagram of the proposed deep EDLSTM for wind power forecasting (The LSTM unit graph in this paper is cited from Ref. [35])

First, the encoder consists of a stack of three-layer LSTMs, which sequentially extracts complex time-dependent features of inputting measured and meteorological data deeply layer by layer with transferring hidden states $\boldsymbol{h}$. And then generate a fixed-length context vector containing the extracted characteristic information. The structure and transmission of information for the decoder are basically identical to those for the encoder. Then, the context vector serves as the initial input to the decoder. Regardless of the updating from the encoder of the context vector, the vector is sent to the first layer of the decoder as its input, and its output is used as the input of the second layer. Sequentially, the third layer output is transformed through the output layer and cyclically fed back to the first layer as its next input. Eventually, the decoder generates a time series of the predicted wind power.

4. Experiments

4.1. Experimental scheme

The scheme of forecasting individual turbine wind power by employing EDLSTM with data augmentation is animatedly illustrated in Fig. 3. Firstly, the measured wind speed and power with the ten-minute resolution are averagely interpolated into, except for the DA1 augmentation measure, data with hourly resolution. All hourly data are segmented into training and testing sets, accounting for 65% and 35%, respectively. Secondly, the measured wind speed and/or wind power data in the training set are separately augmented with the approaches proposed in Section 3.1 to enlarge the data amount to five times the original training set size. i. e., the new data with the four times larger size of the original training set are generated with augmentations. Thirdly, the unexpanded and expanded training sets are individually fed into the benchmark models, i.e., Persistence (PR), simple three-layer backpropagation Neural Networks (NN), basic LSTM RNN (LSTM), Bionic optimized neural networks constructed Adaboost (BA) ensemble leaning (regarded as a popular and advanced hybrid forecasting model [36]), and the proposed deep EDLSTM network to conduct training and obtain multiple learned models. The benchmark models have been introduced in Ref. [37], [38], [39] and their parameters are briefly summarized in Table 2. Finally, the testing set data are imported into the trained models to yield the multistep predicted wind power and to assess and compare the forecasting models' performance.

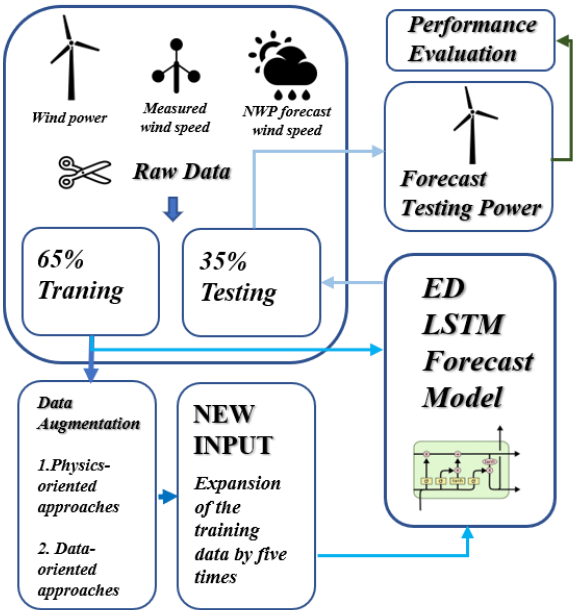


Fig. 3. The main procedure of the data augmentation based EDLSTM for predicting wind power

Table 2. A summary of forecasting models' parameters

Forecasting model	Main parameters
PR	The predicted value for the next moment is the current moment's value.

NN	The input, hidden, and output layers are with 15, 30, and 7 neurons, respectively; sigmoid activation function, and MSE loss function
LSTM	One fully connected dense NN layer, Seven LSTM units, and one dense NN with 7 neurons as output layer; sigmoid activation function, MSE loss function, and Adam algorithm optimizer.
BA	As the performance of a neural network is intimately linked to neuron number in the hidden layer, the genetic algorithm [40], a bionic algorithm, is applied in training iterations to automatically search for the adaptive neuron number and constitute optimized neural networks as Adaboost's base learners. The node number searching interval is set as [10,100] and the max iteration is 50. The Adaboost emphasizes (with bigger weights) data mislearned in the previous base learner to establish an ensemble model that boosts the performance of single base learners. The number of base learners is 10 and Adaboost max iteration is 20.
EDLSTM	As described in Section 3.3 and Fig. 2.

4.2. Data augmentation program

Our data augmentation strategy fine-tunes the data without altering the temporal order of the original data and ensures that the augmented training data and the previous ones maintain statistical consistency. This study augments the training samples and scales up their number to five times the original sample size. The data augmentation techniques explained above, apart from DA5, all involve stochastic perturbation of the original data. Our method is to gradually enlarge the perturbation amplitude and accordingly generate new data four times. For the DA5 method, four new datasets are produced by individually operating autoregressive models based on four machine learning models. Details of the various data augmentation approaches are shown in Table 3.

Table 3. A detailed description of each data augmentation process

Physics-oriented	PA1	The Vestas V90 3 MW wind turbine corresponds to a cut-in and rated wind speed of 4 and 15 m/s, respectively, according to its power curve. Select the measured wind speed v_i in the corresponding interval: $v_i' = v_i + X, X \sim U[-0.1n, 0.1n], n = 1,2,3,4$, where U represents the uniform distribution. Then the power variation corresponding to the wind speed variation is calculated by Eq. (3), and new power data are generated accordingly.
	PA2	The measured wind speeds are randomly substituted with 50 % probability four times with NWP wind speed data with the same timestamps, and the wind power data are added a white noise following N (0,0.1).
	PA3	We select measured wind speeds of the two closest turbines to the target turbine and randomly substitute, with a probability of 15% for each and a total of 30%, the target wind speed dataset. The power data are with the same treatment in PA2.
Data-oriented	DA1	As described in DA1 introduction in Section 3.1.2.
	DA2	Two normally distributed noises, N (0,0.1n) and N (0,0.02n), are separately loaded into the measured wind speed and power data four times, where n=1,2,3,4.
	DA3	As described in DA3 introduction in Section 3.1.2.
	DA4	As described in DA4 introduction in Section 3.1.2.

	DA5	Four learning algorithms to augment measured wind data, such as: $\boldsymbol{x}_t' = f_n(\boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-2}, \boldsymbol{x}_{t-3}, \boldsymbol{x}_{t-4}, \boldsymbol{x}_{t-5}, \boldsymbol{x}_{t-6}), n = 1,2,3,4$, where $\boldsymbol{x}_t'$ is the generating data, $f_i(\boldsymbol{x}_{t-1}, \boldsymbol{x}_{t-2}, \boldsymbol{x}_{t-3}, \boldsymbol{x}_{t-4}, \boldsymbol{x}_{t-5}, \boldsymbol{x}_{t-6})$ represents a single step ahead forecasting model established by learning algorithms. $f_1(.)$ is linear regression, $f_2(.)$ is support vector regression, $f_3(.)$ is classification and regression tree, and $f_4(.)$ is simple three-layer neural networks with 15 hidden neurons regression models, respectively. All four are well-established and widespread machine learning algorithms, and a detailed description of them can be found in Ref. [37] for space constraints.

Note: The units of wind speed and power in the table are m/s and MW, respectively.

4.3. Performance evaluation

Collectively, data-driven wind power forecasting is inherently a matter of using advanced neural networks for regression in which Mean Square Error (MSE) serves as the loss function. So, Root Mean Square Error (RMSE) is naturally selected as the metric to measure the performance of the models. The metric is negative-oriented to the modeling performance, which means a smaller value corresponds to better performance.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{m}(P_i - \widehat{P}_i)^2}{m}} \tag{9}$$

where P_i and $\widehat{P}_i$ are normalized measured and corresponding predicted wind power, m is the sample number of the testing set.

Nevertheless, the RMSE is with a disproportionately big effect of larger errors and, sometimes, is close when comparing some different forecasting models. Therefore, in these cases, Mean Absolute Error (MAE) and Qualification Rate (QR) [41] indices are introduced as below to comprehensively assess the performance of models. MAE uniformly examines the forecasting errors while the QR emphasizes the smaller errors.

$$MAE = \frac{\sum_{i=1}^{m}|P_i - \widehat{P}_i|}{m} \tag{10}$$

$$QR = \frac{1}{m}\sum_{i=1}^{m}\begin{cases}1, \left(1 - \frac{|P_i - \widehat{P}_i|}{Cap}\right) \geqslant Q \\ 0, \left(1 - \frac{|P_i - \widehat{P}_i|}{Cap}\right) < Q\end{cases} \tag{11}$$

where *Cap* is the designed capacity of the turbine. Q is the quantile percentage for qualified predictions, chosen as 90% in this study.

Two statistical tests are employed to check whether there are statistically significant differences exist in the performance of forecasting models. And both of their conference values are set as 0.05. The first is paired T-test for the two comparisons. The null hypothesis H_o: The averages of these samples are equivalent; H_a: The averages are not equivalent. And its test statistic T is:

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{STD_{(\bar{Y}_1 - \bar{Y}_2)}} \sim t_{2l-2} \tag{12}$$

where $\bar{Y}$ is the average and l is the number of samples.

The second is the Friedman test, for multiple comparisons, is harnessed to examine across multiple trials and checks column effects after statistically eliminating potential row effects. [42]

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

H_a: They have a significant difference.

The statistic F is given as:

$$F = \frac{12t}{k(k+1)}\left[\sum_{i=1}^{k} r_i^2 - \frac{k(k+1)^2}{4}\right] \tag{13}$$

where k is the number of columns, r_i is the average value of row i, which follows $\chi^2_{(k-1)}$ distribution under H_o.

5. Results and discussion

This section reveals the experimental results at three levels, firstly, the superiority of the proposed forecasting model is verified by analyzing different models' performance on the original dataset. Secondly, the overall effects of data augmentations on different forecasting algorithms are illustrated by the comparison of their performance before and after data augmentations. And finally, the impacts of various augmentation approaches on the proposed model's forecasting effectiveness are statistically explored.

5.1. Benchmarks and proposed deep EDLSTM model forecasting outcomes

The rescaled measured and NWP wind data of chosen five wind turbines are respectively loaded into the four benchmarks and proposed deep EDLSTM models to make six to twelve hours ahead of wind power forecasts. The RMSE is displayed in Fig. 4. In general, the RMSE of all forecasting models grows as increasing prediction steps. The PR grows faster compared to the other models. The proposed deep EDLSTM outperforms best among all models for multistep power prediction for all wind turbines in almost all cases. The RMSE of the NN, LSTM, BA and EDLSTM all constructed on neural networks is noticeably smaller than the one of PR, suggesting that neural networks can reflect the nonlinear characteristics of wind power. Moreover, these characteristics are better retained by the forecasting models as the networks are deeper and more tailored. On the overall average, the benchmarking PR, NN, LSTM, and BA models have RMSE that is 51.46%, 11.89%, 7.67%, and 4.46% higher than EDLSTM. This demonstrates that the proposed model can efficiently and accurately predict the power generated by the five wind turbines under attention. Besides, EDLSTM's RMSE maintains relative stability with the increasing step, indicating that the seq2seq with multiple inputs and multiple outputs reduces the cumulative error in multistep forecasting. Reasonably, the forecasting algorithms outcome relatively low RMSE of the wind turbines situated on plateau and lakeside, both of which are regarded as flat terrains. In contrast, the unique fjord topography on the Norwegian coast causes wind turbines located on hilltops, valleys, and seasides to be challenging, but handled properly by EDLSTM, to predict their electricity generation. Therefore, the proposed model allows for effective and robust power predictions of wind turbines on several different topographical conditions.

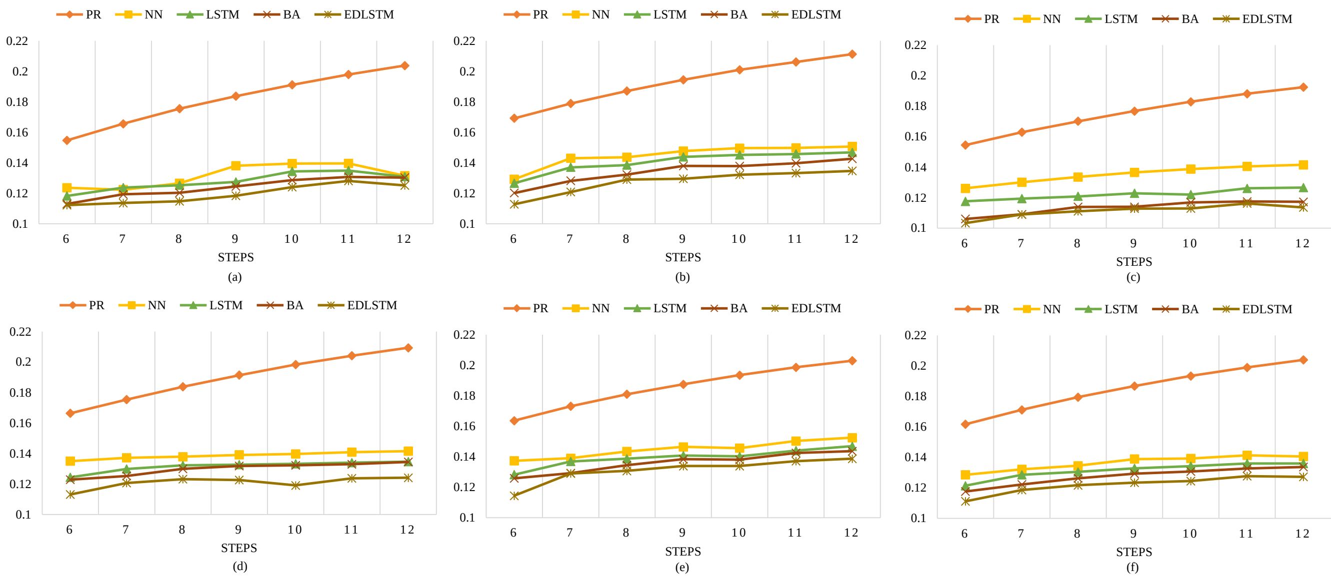


Fig. 4. The multistep performance of benchmarking and deep EDLSTM forecasting models for each turbine: (a) 1.Plateau, (b) 2.Valley, (c) 3.Lakeside, (d) 4.Hilltop, (e) 5.Seaside, (f) Average.

5.2. Holistic validity of data augmentations

Aiming to investigate the applicability of data augmentation in wind power prediction, the original measured data are enlarged following the eight augmentation approaches presented in Section 3.1 and are predicted by the four benchmarks and the proposed EDLSTM models. The RMSE for the six to twelve-step forecasts by the forecasting algorithms based on the eight data-augmented sets is averaged separately. The results are compared to the RMSE equally averaged of the models without augmentations. Fig. 5 shows the comparison, and Table 4 offers their performance difference with paired T-test.

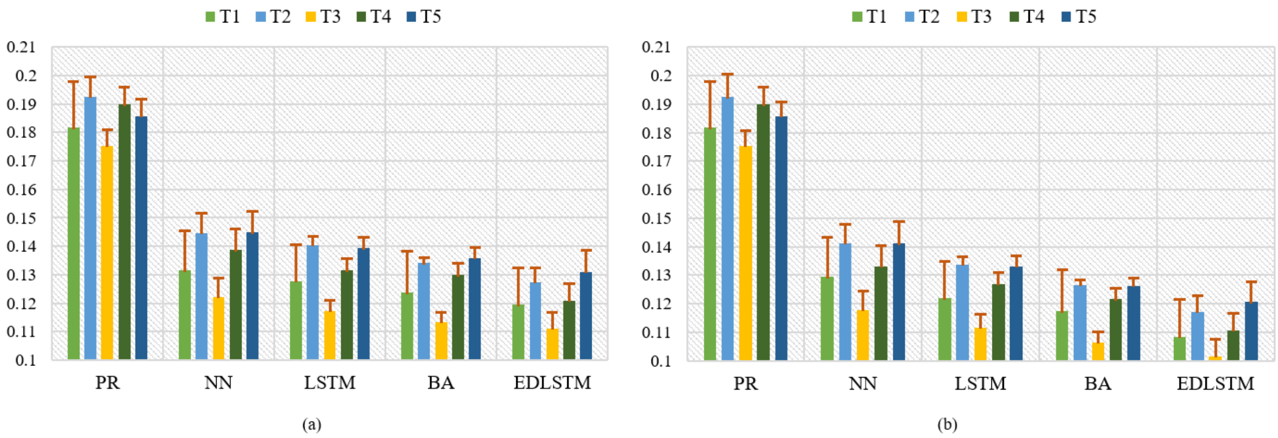


Fig. 5. The overall average RMSE of multistep forecasting models without and with data augmentations for each turbine: (a) Without augmentations, (b) With augmentations.

Table 4. The *T* test for average performance difference between without and with data augmentations

Paired *T*	**PR**	**NN**	**LSTM**	**BA**	**EDLSTM**
Mean	0	0.003876	0.005813	0.004610	0.010291
***p*-values**	/	0.002219	0.000083	0.000173	0.000001

As can be seen, the average effect of data augmentation is tightly linked to forecasting algorithms. The RMSE of PR with data augmentation is the same as the previous one for all wind turbines in focus. The reason is there is no learning process in the PR method and its RMSE remains the same when the used data augmentations give stochastic perturbations in data or generations of new data based on patterns of primitive data. So, it is meaningless to further discuss the augmentation in the PR approach. Within one STD, there is an apparent difference, with p-values smaller than 0.05, between RMSE of all network-based NN, LSTM, BA, and EDLSTM forecasting algorithms. It can be interpreted that these algorithms can not only respectively learn the dominant or trending patterns in the input space, but data augmentations also provide additional valuable information in these network-based models training phrases.

Most notably, a significant improvement, with a statistical average difference over 0.0102, in the performance of the EDLSTM forecasting algorithm is evident with augmented input data. On the one hand, it means that the limited original data restrict the proposed deep learning model's potential or possibly cause overfitting. On the other hand, it demonstrates that the augmented data more adequately train the complex deep networks to yield better predictions by insight into more hidden and sophisticated patterns in the forecasting. In addition, the STD of RMSE between multiple predictions shows no significant variation before and after data augmentations, which points out that the effects of data augmentations are approximate for each step. Generally, the average RMSE of augmented models of NN, LSTM, and BA separately grows by 21.47%, 13.30%, and 7.60% compared with augmented EDLSTM.

To more explicitly show outcomes of the various data-augmented models, the RMSE of each step prediction based on the eight augmentation approaches is averaged and plotted in Fig. 6.

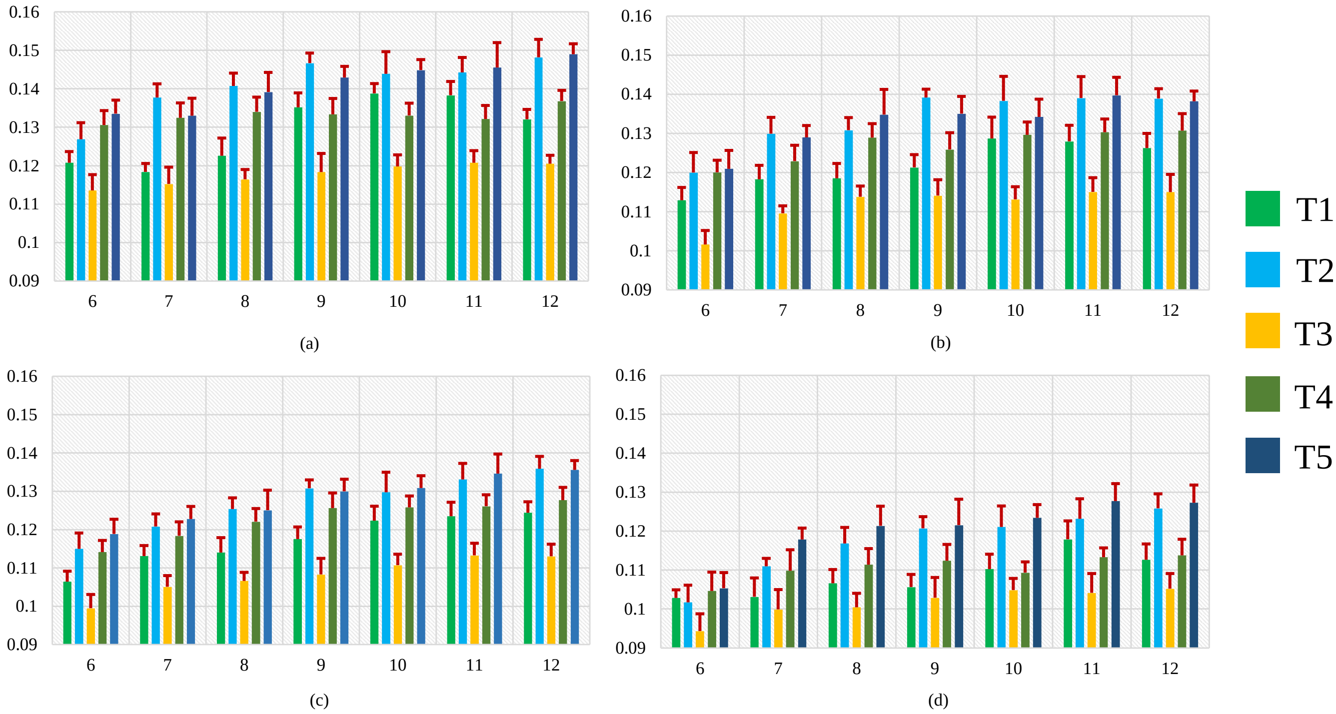


Fig. 6. The multistep average RMSE of forecasting models with data augmentations for each turbine: (a) NN, (b) LSTM, (c) BA, (d) EDLSTM.

By comparing Fig. 6 with Fig. 4, it can be found that: first, the tendency of gradually increasing RMSE persists of data-augmented multistep predictions. Secondly, the augmented EDLSTM model outperforms its counterpart based on raw data in almost every step of prediction for all wind turbines. And thirdly, the power prediction of T3 wind turbine is the best, corresponding to the RMSE of the data augmented EDLSTM model is barely less than 0.11, and the second-best one is T1. Furthermore, the predictions for T2, T4, and T5, located in complex terrain, are also significantly improved. Thus, data augmentation improves EDLSTM for power forecasting, resulting in satisfactory reductions in model RMSE errors.

5.3. Competition between diverse data augmentation methodologies

The superiority of data augmentation approaches as a whole in wind power prediction is elaborated in Section 5.2. To further investigate which data augmentation approaches are more effective, the average and STD of RMSE for each step of prediction by algorithms based on different augmentation approaches are taken and presented in Fig. 7. As can be seen, there is no obvious regularity in the average multistep forecasting performance with different augmentation-based models. That is, the results of various augmentation approaches in different forecasting algorithms are not tendentious. The overall RMSE of distinct augmentations is comparable in NN, LSTM, and BA but the opposite is the view in EDLSTM. Nevertheless, certain patterns exist for augmentations in the prediction of different turbines. Regardless of what augmentations, the errors in predictions for turbines in flatter terrain are smaller, consistent with the predictions without augmentations.

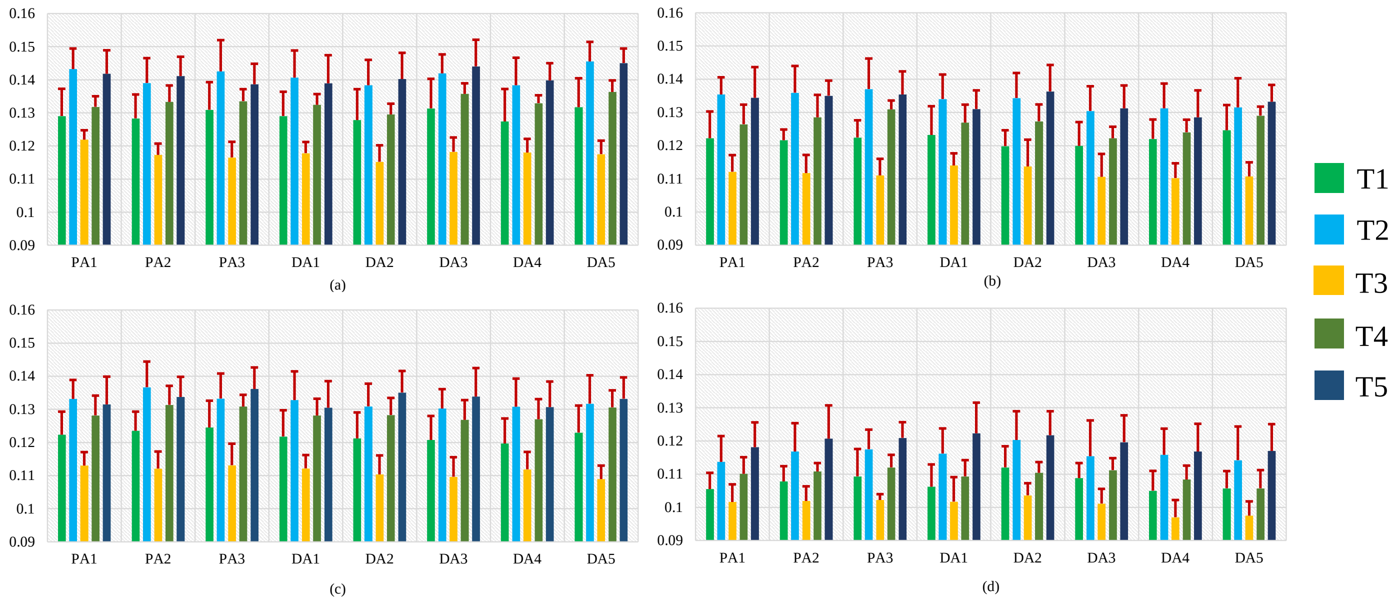


Fig. 7. The multistep average RMSE of forecasting models with various data augmentations for each turbine: (a) NN, (b) LSTM, (c) BA, (d) EDLSTM.

As a further statistical examination to test the variation in different data augmentation in multistep predictions, the Friedman test to answer whether there is a difference between the RMSE averages of the five wind turbines with different augmentations in the same time step. The p-values are demonstrated in Table 5. Among the power forecasts based on data augmentations for all turbines, The effect of different augmentation approaches for forecasting models is not statistically significant in most cases, such as in NN, LSTM, and most cases of BA. Particularly, the proposed EDLSTM models' RMSE, with a relatively complex p-value set, differs only in sixth and seventh step forecasts with varying augmentations. Additionally, in view of the EDLSTM's favorable outperformance in wind power forecasting, the decrease rate of average multistep RMSE for each augmented versus unaugmented model based on the same forecasting algorithm is computed. The rate is averaged among five turbines and illustrated in Fig. 8. The p-value for the multivariate comparison between these RMSE decrease rates is 0.00033, much less than 0.05, indicating that overall improvements in EDLSTM performance with various augmentations are statistically different. In general, based on RMSE, PA3, PA2, DA1, and PA1 provide modest improvements, from 7.87% to 9.96%, to the EDLSTM model, while DA5, DA4, DA3, and DA2 improve, sequentially from 10.80% to 11.36%, the model relatively substantially.

Table 5. The p-values of RMSE Friedman test within five turbines for multiple comparisons in different data-augmented approaches

***P*-values**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
NN	0.4717	0.2772	0.1013	0.54	0.1705	0.1046	0.7608
LSTM	0.1274	0.6809	0.3268	0.6809	0.0183	0.0335	0.6113
BA	*0.0202*	0.4084	0.3445	0.0558	0.1775	0.532	0.1507
EDLSTM	*0.0049*	*0.0012*	0.0626	0.3041	0.1213	0.0901	0.1239

Note: The p-values less than 0.05 are marked in italics meaning H_0 is rejected.

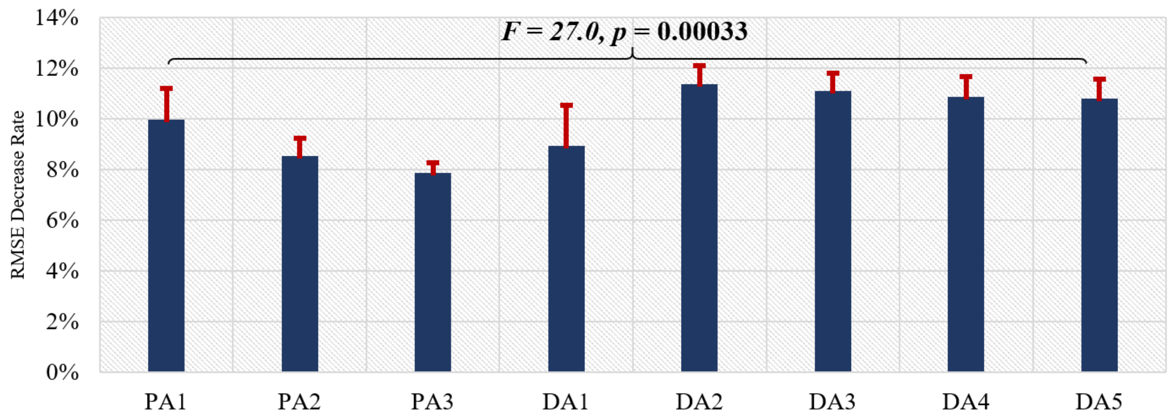


Fig. 8. The average RMSE decrease rate of multistep EDLSTM forecast with various augmentations for averaging five turbines

Despite the varying decrease degrees in RMSE for the EDLSTM models with different augmentation approaches, the difference is minimal between some approaches, like DA4 and DA5. To further compare the effects of different augmentations, the average MAE and QR90 of forecasts with the same scenario as in Fig. 8 are gained and their change rates before and after augmentations are calculated and tested in Fig. 9 and 10. The p-value of MAE decrease rate comparison is 0.0023, less than 0.05, also smaller than its counterpart of RMSE, which also means varying augmentations give statistically different boosts in EDLSTM. Similar to Fig

8, the DAs are better than PAs, but Fig. 9 offers a clearer distinction between several DAs. DA4 and DA5 have a greater MAE decline, 8.97% and 8.82%, than DA2 and DA3, 8.49% and 7.79%, which generally indicates that the former two provide closer predictions to the real values. But DA4 and DA5 may have big deviations in some forecasting points, so these data-oriented augmentations are quite close in Fig. 8. The *p*-value of QR90 increase rate comparison is 0.0052, bigger than 0.05, which illustrates different augmentations have no significant different improvements, around 12% to 13%, in QR90. This phenomenon reveals that either augmentation technique can elevate the qualification rate of the EDLSTM model in a relatively similar amount and provide satisfactory forecasts in terms of this evaluation index.

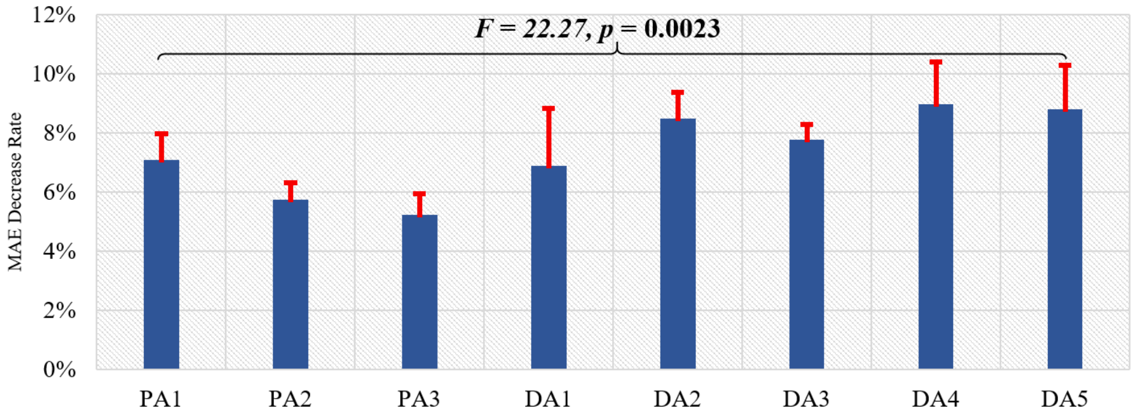


Fig. 9. The average MAE decrease rate of multistep EDLSTM forecast with various augmentations for averaging five turbines

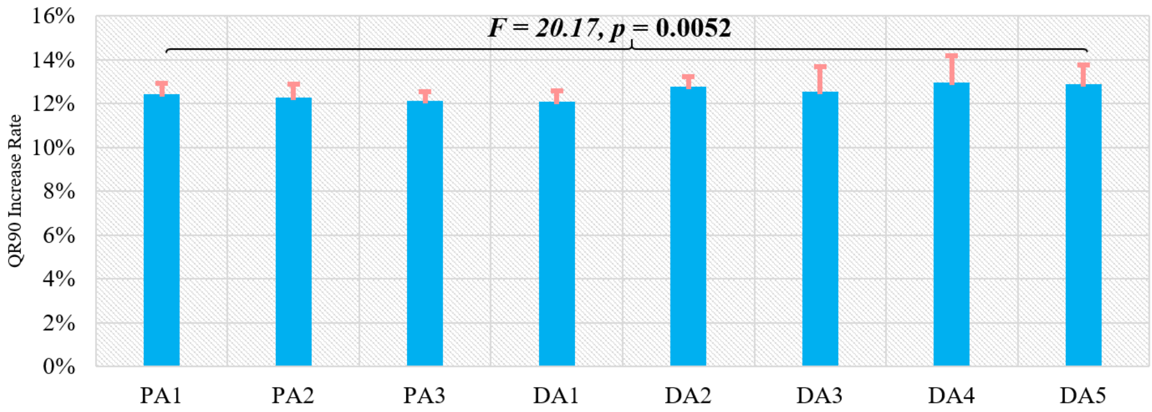


Fig. 10. The average QR90 increase rate of multistep EDLSTM forecast with various augmentations for averaging five turbines

To summarize, the impact of the different data augmentation methods on the benchmark models is not significantly different. However, the improvement for the deep EDLSTM is slightly varied, unremarkable in QR90 metric. DAs, on the whole, outperform PAs in RMSE and MAE, and MAE further reveals that DA4 and DA5 have edges in the DA methods.

6. Conclusions

This paper initially scrutinizes the usefulness of data augmentation approaches in wind power forecasting and proposes a multi-input and multi-output prediction algorithm with verified superiority. Inferences on the results of multistep forecasting five wind turbines with various topologies, conclusions are given as follows:

The proposed seq2seq-based deep EDLSTM enables highly effective and robust multistep power forecasting, by highlighting the sequential dependence of the problem, for wind turbines under different terrain conditions. Also, compared with the benchmark PR, NN, LSTM, and BA algorithms, its overall RMSE is lowered by 33.89%, 10.60%, 7.12%, and 4.27%, respectively.

Since EDLSTM is a complex deep learning model, its strength requires so-called big data. It is demonstrated that five-fold expansions of the primary data with data augmentations statistically boost neural network-based NN, LSTM, BA, and EDLSTM wind power forecasting capabilities. The boost is particularly evident in EDLSTM, where, on average, the performance of the data-augmented model provides better forecasting with lower RMSE, which is 10.2% smaller than its counterpart without data augmentation. This boosting can be interpreted as expanding the training set, it is equivalent to adding a regular term to the loss function when training models, which can effectively avoid overfitting. Besides, due to the stochasticity involved in data augmentations, the learned model built on the techniques presents better robustness. Moreover, the data-augmented EDLSTM edges over the benchmarks, PR, NN, LSTM, and BA with the same expanding inputs, extending to 40.63%, 17.67%, 11.74%, and 7.06% decrease in RMSE, respectively since the proposed EDLSTM further learned deeper information, like signal decompositions, of the wind data by mentioned augmentation techniques.

The impact of the eight data augmentation approaches employed, three physics-oriented and five data-oriented, on wind power prediction is forecasting arithmetic sensitive. For the proposed well-performing EDLSTM, various augmentations can approximately, by over 12%, boost the forecasting qualification rate at the 90% threshold. But augmentations improve the forecasting performance to slightly different degrees when evaluated by RMSE and MAE: multistep and multiturbine meanly, the improvement varies from approximately 7.87% to 11.36% of RMSE and 5.24% to 8.97% of MAE within one standard deviation, and generally, data-oriented augmentations outperform physics-oriented ones. Among data-oriented augmentations, the results illustrate that EDLSTM's forecasting RMSE is significantly decreased even by simply appending noisy and randomly perturbing, or moving data the same way as sophisticated statistical data decomposition and learning data generation, however, as per MAE, the latter two provide overall closer predictions to the real power.

Our future research, on the basis of this paper, foresees to further investigate de facto more advanced data augmentation techniques and integrate them into the proposed model to conduct in-depth point and probability predictions and attempt industrial applications in extensive comparisons with other forecasting models.

Additionally, ensuing policy recommendations may be extrapolated:

Drawing on state-of-the-art deep learning techniques and increasing computational abilities, wind power forecasting and deriving data issues in energy fields shall be approached progressively from traditional statistical and parameters-sensitive classical machine learning methods to deep learning approaches that can automatically identify complex patterns. Besides, the sophisticated deep networks are particularly reliant on data amounts. Motivated by this article, limited data of wind parks or other energy sectors could be artificially enlarged by appropriate data augmentations to serve as the steppingstone for further applications of deep learning to challenge related scientific and engineering difficulties.

References

[1] T. M. Letcher, *Wind energy engineering: A handbook for onshore and offshore wind turbines*. Academic Press, 2017.

[2] E. Rahimi, A. Rabiee, J. Aghaei, K. M. Muttaqi, and A. E. Nezhad, "On the management of wind power intermittency," *Renewable and Sustainable Energy Reviews,* vol. 28, pp. 643-653, 2013.

[3] Y. Cai and F.-M. Bréon, "Wind power potential and intermittency issues in the context of climate change," *Energy Conversion and Management,* vol. 240, p. 114276, 2021.

[4] H. Liu, C. Chen, X. Lv, X. Wu, and M. Liu, "Deterministic wind energy forecasting: A review of intelligent predictors and auxiliary methods," *Energy Conversion and Management,* vol. 195, pp. 328-345, 2019.

[5] S. Hanifi, X. Liu, Z. Lin, and S. Lotfian, "A critical review of wind power forecasting methods—past, present and future," *Energies,* vol. 13, no. 15, p. 3764, 2020.

[6] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science,* vol. 313, no. 5786, pp. 504-507, 2006.

[7] P. Kim, "Matlab deep learning," *With machine learning, neural networks and artificial intelligence,* vol. 130, p. 21, 2017.

[8] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, "A review of deep learning for renewable energy forecasting," *Energy Conversion and Management,* vol. 198, p. 111799, 2019.

[9] C. Sun, A. Shrivastava, S. Singh, and A. Gupta, "Revisiting unreasonable effectiveness of data in deep learning era," in *Proceedings of the IEEE international conference on computer vision,* 2017, pp. 843-852.

[10] X. Wang, Y. Shi, and K. M. Kitani, "Deep supervised hashing with triplet labels," in *Asian conference on computer vision*, 2016: Springer, pp. 70-84.

[11] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data,* vol. 6, no. 1, pp. 1-48, 2019.

[12] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *arXiv preprint arXiv:1409.3215,* 2014.

[13] Y. Zhang, Y. Li, and G. Zhang, "Short-term wind power forecasting approach based on Seq2Seq model using NWP data," *Energy,* p. 118371, 2020.

[14] M. Pirhooshyaran, K. Scheinberg, and L. V. Snyder, "Feature engineering and forecasting via derivative-free optimization and ensemble of sequence-to-sequence networks with applications in renewable energy," *Energy,* vol. 196, p. 117136, 2020.

[15] K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078,* 2014.

[16] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning* (no. 2). MIT press Cambridge, 2016.

[17] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "Autoaugment: Learning augmentation strategies from data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 113-123.

[18] T. DeVries and G. W. Taylor, "Dataset augmentation in feature space," *arXiv preprint arXiv:1702.05538,* 2017.

[19] X. Deng, H. Shao, C. Hu, D. Jiang, and Y. Jiang, "Wind power forecasting methods based on deep learning: A survey," *Computer Modeling in Engineering & Sciences,* vol. 122, no. 1, pp. 273-302, 2020.

[20] C. Yildiz, H. Acikgoz, D. Korkmaz, and U. Budak, "An improved residual-based convolutional neural network for very short-term wind power forecasting," *Energy Conversion and Management,* vol. 228, p. 113731, 2021.

[21] A. Kisvari, Z. Lin, and X. Liu, "Wind power forecasting–A data-driven method along with gated recurrent neural network," *Renewable Energy,* vol. 163, pp. 1895-1909, 2021.

[22] F. Shahid, A. Zameer, and M. Muneeb, "A novel genetic LSTM model for wind power forecast," *Energy,* vol. 223, p. 120069, 2021.

[23] G. Memarzadeh and F. Keynia, "A new short-term wind speed forecasting method based on fine-tuned LSTM neural network and optimal input sets," *Energy Conversion and Management,* vol. 213, p. 112824, 2020.

[24] M. H. Lipu *et al.*, "Artificial Intelligence Based Hybrid Forecasting Approaches for Wind Power Generation: Progress, Challenges and Prospects," *IEEE Access,* 2021.

[25] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, and N. Carvalhais, "Deep learning and process understanding for data-driven Earth system science," *Nature,* vol. 566, no. 7743, pp. 195-204, 2019.

[26] G. Giebel and G. Kariniotakis, "Wind power forecasting—A review of the state of the art," *Renewable energy forecasting,* pp. 59-109, 2017.

[27] A. Ng, "Advice for applying machine learning," in *Machine learning*, 2011.

[28] H. Chen, Y. Birkelund, S. N. Anfinsen, R. Staupe-Delgado, and F. Yuan, "Assessing probabilistic modelling for wind speed from numerical weather prediction model and observation in the Arctic," *Scientific Reports,* vol. 11, no. 1, pp. 1-11, 2021.

[29] Q. Dong, Y. Sun, and P. Li, "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," *Renewable Energy,* vol. 102, pp. 241-257, 2017.

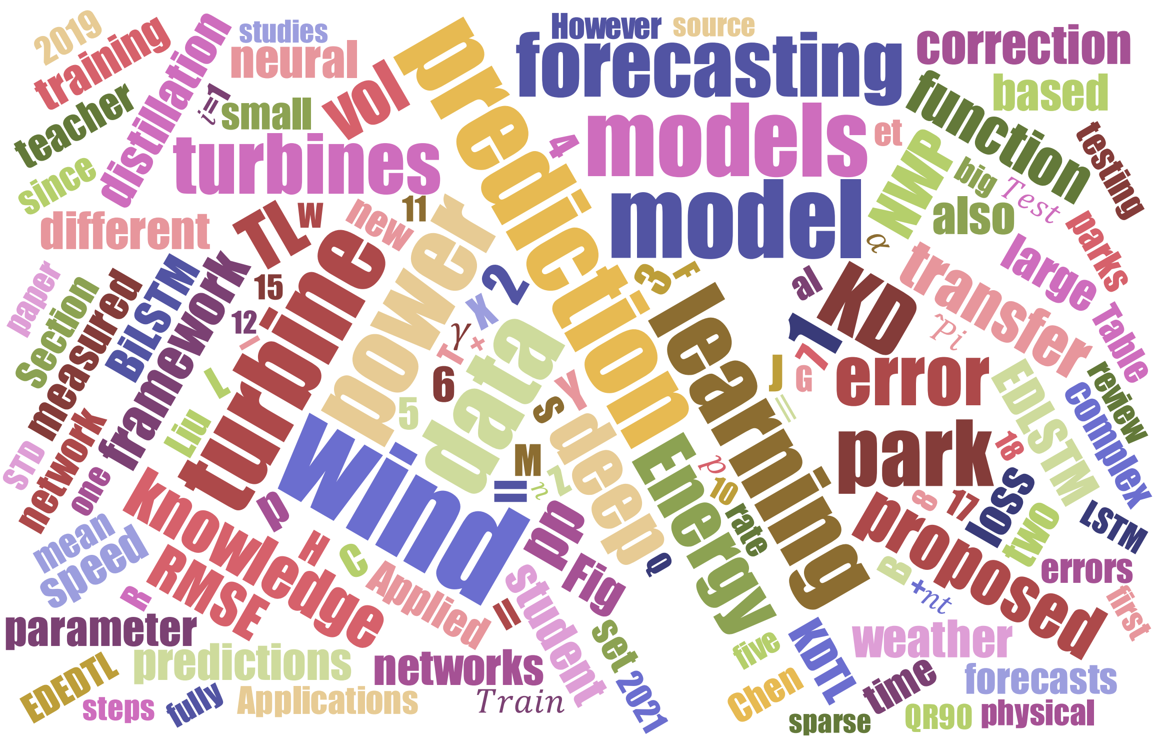
[30] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, "STL: A seasonal-trend decomposition," *Journal of official statistics,* vol. 6, no. 1, pp. 3-73, 1990.

[31] J. Li, J. Zhou, and B. Chen, "Review of wind power scenario generation methods for optimal operation of renewable energy systems," *Applied Energy,* vol. 280, p. 115992, 2020.
[32] H. Chen, Y. Birkelund, S. N. Anfinsen, and F. Yuan, "Comparative study of data-driven short-term wind power forecasting approaches for the Norwegian Arctic region," *Journal of Renewable and Sustainable Energy,* vol. 13, no. 2, p. 023314, 2021/03/01 2021, doi: 10.1063/5.0038429.
[33] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," 1999.
[34] R. C. Staudemeyer and E. R. Morris, "Understanding LSTM--a tutorial into Long Short-Term Memory Recurrent Neural Networks," *arXiv preprint arXiv:1909.09586,* 2019.
[35] C. Olah, "Understanding lstm networks," 2015.
[36] Z. Qian, Y. Pei, H. Zareipour, and N. Chen, "A review and discussion of decomposition-based hybrid models for wind energy forecasting applications," *Applied energy,* vol. 235, pp. 939-953, 2019.
[37] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.
[38] J. Lee, W. Wang, F. Harrou, and Y. Sun, "Wind power prediction using ensemble learning-based models," *IEEE Access,* vol. 8, pp. 61517-61527, 2020.
[39] W. Sun, T. Zhang, R. Tao, and A. Wang, "Short-Term Photovoltaic Power Prediction Modeling Based on AdaBoost Algorithm and Elman," in *2020 10th International Conference on Power and Energy Systems (ICPES)*, 2020: IEEE, pp. 184-188.
[40] D. Whitley, "A genetic algorithm tutorial," *Statistics and computing,* vol. 4, no. 2, pp. 65-85, 1994.
[41] L. Lijuan, L. Hongliang, W. Jun, and B. Hai, "A novel model for wind power forecasting based on Markov residual correction," in *IREC2015 The Sixth International Renewable Energy Congress*, 2015: IEEE, pp. 1-5.
[42] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference: Revised and Expanded*. CRC press, 2014.

Paper V.

Knowledge distillation with error-correcting transfer learning for wind power prediction

https://doi.org/10.13140/RG.2.2.12410.57286



Knowledge distillation with error-correcting transfer learning for wind power prediction (Preprint)

Hao Chen[a,c,*], Yngve Birkelund[b,c]

* = corresponding author, hao.chen@uit.no

[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

Abstract

Wind power prediction, especially for turbines, is vital for the operation, controllability, and economy of electricity companies. Hybrid methodologies combining advanced data science with weather forecasting have been incrementally applied to the predictions. Nevertheless, individually modeling massive turbines from scratch and downscaling weather forecasts to turbine size are neither easy nor economical. Aiming at it, this paper proposes a novel framework with mathematical underpinnings for turbine power prediction. This framework is the first time to incorporate knowledge distillation into energy forecasting, enabling accurate and economical constructions of turbine models by learning knowledge from the well-established park model. Besides, park-scale weather forecasts non-explicitly are mapped to turbines by transfer learning of predicted power errors, achieving model correction for better performance. The proposed framework is deployed on five turbines featuring various terrains in an Arctic wind park, the results are evaluated against the competitors of ablation investigation. The major findings reveal that the proposed framework, developed on favorable knowledge distillation and transfer learning parameters tuning, yields performance boosts from 3.3 % to 23.9 % over its competitors. This advantage also exists in terms of wind energy physics and computing efficiency, which are verified by the prediction quality rate and calculation time.

Keywords: Renewable energy prediction; Knowledge distillation; Deep learning; Error correction; Transfer learning; Arctic

Abbreviations

NWP	Numerical Weather Prediction
TL	Transfer learning
LSTM	Long Short-Term Memory
EDLSTM or ED	Encoder-Decoder Long Short-Term Memory neural networks
KD	Knowledge Distillation
MMD	Maximum Mean Discrepancy
MSE	Mean Square Error
Bi-LSTM	Bidirectional LSTM
RMSE	Root Mean Square Error
QR90	Qualification Rate at the 90% threshold
GeoMean	Geometric mean
STD	Standard deviation
T#.	Wind turbines with different terrain – turbine number

1. Introduction

The exploitation of renewable energy is a propitious approach to achieving carbon neutrality. Wind energy, as one of the foremost renewable energy sources, has gained increasing prominence in various countries for its abundant availability, technological maturity, and favorable financial support [1]. However, the volatility and randomness of the natural wind, especially when sited in complex terrain conditions [2], create tremendous uncertainty in wind power generation. The volatile electricity generated by wind turbines causes adverse effects on power quality, grid dispatch, and the stability and security of power system operation [3]. Therefore, as larger-scale wind power is integrated into the grid, there is a considerable necessity to further develop associated energy forecasting strategies [4] to reduce the impact of renewable on the grid and improve energy efficiency.

Wind power prediction can be categorized into physical, statistical, and hybrid approaches [5]. The first is appropriate for relatively long-term (days) forecasts through atmospheric physics modeling called. The second is suitable for relatively short-term (minutes or hours) predictions. However, the current prediction investigations, which usually combine the above two methods and can extrapolate the prediction-time scenario to realize higher accuracy, are attracting more interest [6].

1.1. Previous work

Extensive research has been conducted using Numerical Weather Prediction (NWP) downscaling [7], [8], signal decomposition [4], statistical regression [9], machine learning [5], [10], deep learning [11], and more for wind power prediction and has achieved satisfactory results in many wind park experiments.

The state-of-the-art of employing deep learning associated approaches to predict wind power is sufficiently summarized and reviewed in [11], [12], [13]. These studies are primarily devoted to the decomposition of wind energy-related sequences using different data processing techniques and developing advanced deep learning algorithms to recognize the in-depth features of these sequences. Since wind power is primarily driven by weather, adding information from suitable weather forecasts to a generation prediction mode can lead to an accuracy boost [14]. However, it is common to downscale mesoscale NWP (wind park-scale or larger) applying meteorological [8] or statistical methodologies [15], [16] or to input these data directly into the prediction model for automatic recognition [17], [18]. The former involves substantial physical assumptions and computational time whereas the latter loses differences between turbines in the same wind park.

As the availability of wind energy data and the accuracy of power prediction increases, so does the complexity of the forecasting models. Since different researchers have developed various forecasting models for the individual park, when faced with the problem of a new site, it is required to develop a fresh model from scratch.

Transfer learning (TL) is a machine learning paradigm that applies some kind of learning algorithms to derive information from one or more application scenarios and thereby contribute to enhancing the learning performance for the target scenarios [19]. There appear to be a few efforts in recent years to apply TL to wind power prediction. Hu Q et al. [20] tried to transfer wind speed information from a data-rich old operating wind park to a newly-built park and successfully used stacked denoising autoencoders to forecast the speed with each other data for four wind parks in the north of China. Qureshi AS et al. [21] speeded up and optimized the deep sparse autoencoder power prediction model learning by only training one park data and transferring the gained knowledge to others with sounds weight initializations. Yin H et al. [22] focused on multiple wind parks power forecasting with TL in Inner Mongolia and applied CNNs-LSTM to extract sequence features of source parks and deliver the features to build a good model for the target wind farm. These studies all directly and successfully transferred information from other wind farms to the intended one, but without accounting for the forecasting issue for turbines. Liu X et al. [23] considered turbines prediction by a proposed a deep and TL framework with the turbine Supervisory Control And Data Acquisition (SCADA) data, and by processing the homogeneity and heterogeneity among different wind turbines, they realized the ultra-short-term power prediction within one hour with high accuracy. However, the study is a purely statistical and short-term forecast; to achieve the long-term prediction, weather forecasts must be factored.

The forecasting error includes unlearned information, analyzed in [24], which can be extracted as a complementary for park power prediction models. A few studies consider the magnitude of prediction errors. Ding M et al. [25] used a gated recurrent unit neural network to correct the NWP wind speed error and applied the corrected speed to model the power curve with efficiency. Sun Z et al. [26] directly modeled the error sequence trend of preliminary prediction results and combined the predicted error to obtain the final predictions. However, most of these studies treated the prediction errors as separate time series, which are not always reliable because the error series may be very white-noisy and hard to be forecasted or the accuracy of their prediction decreases rapidly with the forecasting time step [27].

1.2. Contributions

As our previous work published *Data-augmented sequential deep learning for wind power* forecasting and its literature showed that the work returned to the physical process of wind power generation, the statistical characteristics of wind data, and the nature of deep learning to approach the prediction problem for the turbine in varying terrain conditions [17] and yielded a delicate forecasting structure promising multistep prediction.

However, there are still the following points where enhancements are possible according to our further investigations.

1. Deep learning-based models require multiple layers of uniquely designed neural network structures to realize the intrinsic features of the data [28]. And training large deep networks is very time-consuming and computationally intensive. Therefore, it is proposed that certain pre-training techniques, such as Knowledge Distillation (KD) [29], can be adopted to *distill* useful information, knowledge, from the large pre-trained, whole park, *teacher* model and *condense* it into smaller scale turbine prediction, *student*, models.
2. The NWP information in hybrid prediction models is usually for the whole wind park and does not precisely reflect the individual turbine's future meteorological condition. Besides, the wake effect (wind is strongly perturbed, decreased kinetic energy and added turbulence behind blades of a turbine), together with the turbulence induced by the micro-scale topography in complex terrain, thereby rendering further forecasting difficulties. Typically, single wind turbine meteorological modeling considers NWP results and simulates turbine wind conditions with Computational Fluid Dynamics

(CFD) [30], [31]. This paper will bypass this complex approach based on multiple physical assumptions and indirectly integrate the meteorological information of turbines into the whole prediction model by data science.

3. Wind power forecasting is essentially reducible to a regression problem, so regression diagnostics in statistics, especially error analysis and correction, could be incorporated into the prediction model. Hence, this paper achieves the detection and forecasting of prediction errors through advanced deep learning approaches. Moreover, weather information is ingeniously embedded into the final prediction model by the errors correcting.

The presented paper fully addresses the above concerns by innovatively developing a framework for predicting wind turbine power with solid mathematical derivation. It is the first time, through KD, to deploy a large sequential deep learning forecasting model with multiple inputs and outputs for big data in wind parks on a fast-running small-scale turbine forecasting model; the framework fully exploits the prediction error value and ingeniously downscales the NWP information corresponding to wind parks non-explicitly to the turbine scale by transfer learning and primitive and inverse function transformation. The effectiveness and quickness of the framework are experimentally verified on wind turbines in different terrains. Furthermore, the framework has extensive applicability in other fields since it does not involve specific energy-physics and geographical factors.

The rest of this article is organized as follows. Section 2 elaborates on the deep learning techniques involved in the proposed wind power prediction framework. Section 3 describes case study experiment data and vividly deduces the mathematical principles underlying the framework. In section 4, the experimental procedure and evaluation are briefly stated. The experimental results for comprehensive verification of the framework are thoroughly discussed in section 5. Finally, the main findings and policy recommendations are exhibited in Section 6.

2. Methodology

This section elaborates on the peripheral techniques utilized in this paper, starting with Transfer Learning (TL), followed by Knowledge Distillation (KD), an important TL technique, and finally, prediction algorithms and their resultant error corrections.

2.1. Transfer learning

An essential challenge in employing machine learning applications in engineering is the weak displacement of models, i.e., existing models do not transfer well into new areas. Firstly, Many machine learning application applications are trained with small data: traditional learning algorithms tend to suffer from overfitting problems due to small data size [32]. Thus, the established models cannot be well extended to new scenarios. Secondly, strong robustness is required for machine learning models: classical algorithms assume that training and testing data are from the same statistical distribution [33], which are sometimes unrealistic in engineering. Finally, personalization and data security; in largescale practical applications, where datasets often belong to multiple owners and cannot be disclosed to each other for security reasons. Therefore, the learning model should extract the intrinsic of each dataset and transfer to new scenarios [34].

The emerging TL responds to this problem. It is a methodology that addresses data distributions in the source domain (where the model has been trained) and the target domain (where the trained model will be implemented), which are similar but not identical, to promote efficiency in machine learning [35]. The core in a sound TL is finding the similarity between the mentioned two scenarios to achieve a so-called adaptive learning. TL is categorized into two types with the based on the features in the two scenarios: homogeneous and heterogeneous [36]. The former is only considered because the data used are only wind sequence data and designed transfer process is elaborate. TL, especially deep TL, directly improves performance on different tasks by recognizing patterns directly on the original data and transferring the models to other original data. It enables the automagical extraction on more expressive features and meets the end-to-end demands of real-world applications.

In the present study, TL architecture is established with the development of an adaptive sparse deep learning network for univariate pattern recognition. The structure is shown in Fig. 1.

In fact, adaptive learning is effectively applied by identifying and reducing differences between the source and target domains through simple and executable transformations to transfer learned models from source domains to target ones. The Kullback Leibler Divergence (D_{KL}), also named relative entropy in (1), is mostly employed to identify similarities between two distributions $p(x)$ and $q(x)$. Practically, Practically, since D_{KL} involves integration and logarithmic operations, a similar metric Maximum Mean Discrepancy (MMD), in (2), as an approximation of bivariate divergence is available in the [37].

$$D_{KL}(p(x) \parallel q(x)) = \int_{-\infty}^{\infty} p(x) \ln \frac{p(x)}{q(x)} dx \tag{1}$$

$$MMD(x, x_t) = sup\left(\frac{1}{N}\sum_{i=1}^{N} f(x^i) - \frac{1}{N_t}\sum_{j=1}^{N_t} f(x_t^j)\right) = \left[\frac{1}{N^2}\sum_{i,j=1}^{N} f(x^i, x^i) - \frac{2}{N \cdot N_t}\sum_{i,j=1}^{NN_t} f(x^i, x_t^j) + \frac{1}{N_t^2}\sum_{i,j=1}^{N_t} f(x_t^i, x_t^j)\right]^{\frac{1}{2}} \tag{2}$$

where $sup(.)$ is supremum; x^i and $x_t{}^j$ are the i-th and j-th sample in the source and target domains, respectively; $f(.)$ donates feature mapping function in TL; N and N_t represents source and target domain size.

For deep learning network (extensively literature detailing DL [28]). The utilized sparse DL's (shorten as $D1$) loss function L_{D1} is expressed as follows:

$$L_{D1} = \frac{1}{n}\sum_{i=1}^{n} \left(X_i - \hat{X}_i\right)^2 + \beta\Omega_w + \delta\Omega_s \tag{3}$$

where $\frac{1}{n}\sum_{i=1}^{n} \left(X_i - \hat{X}_i\right)^2$ is the Mean Squared Error (MSE), typical loss function for regression task, $\Omega_w = \frac{1}{2}\sum\left(W_{i,j}\right)^2$ donates the L_2 regularization term (handling over fitting) of NN weights, $\Omega_s = D_{KL}(p_i \parallel \hat{p}_i)$ represents the sparse regularization term that enforces constraints of sparsity in outputs from hidden layers to minimize the calculated data distribution $\hat{p}_i$ and inputs actual distribution p_i.

The well-learned model is normally can be directly transferred to process new datasets. Since the old and new datasets are merely similar but not identical, an adaptive MMD-based mechanism (an Optimizer named MMD comparator) is introduced into the model transfer to further improve the new model's ($D2$) accuracy. The loss function L_{D2} of $D2$:

$$L_{D2} = L_{D1} + \gamma\Omega_{MMD} \tag{4}$$

where $\Omega_{MMD} = \sum_{layers} MMD(X_i, X_i')$ is the divergence regularization term to minimize the gaps between the layer to layer of old and new models. So, L_{D2} optimized the transferred old model to fit new data better.

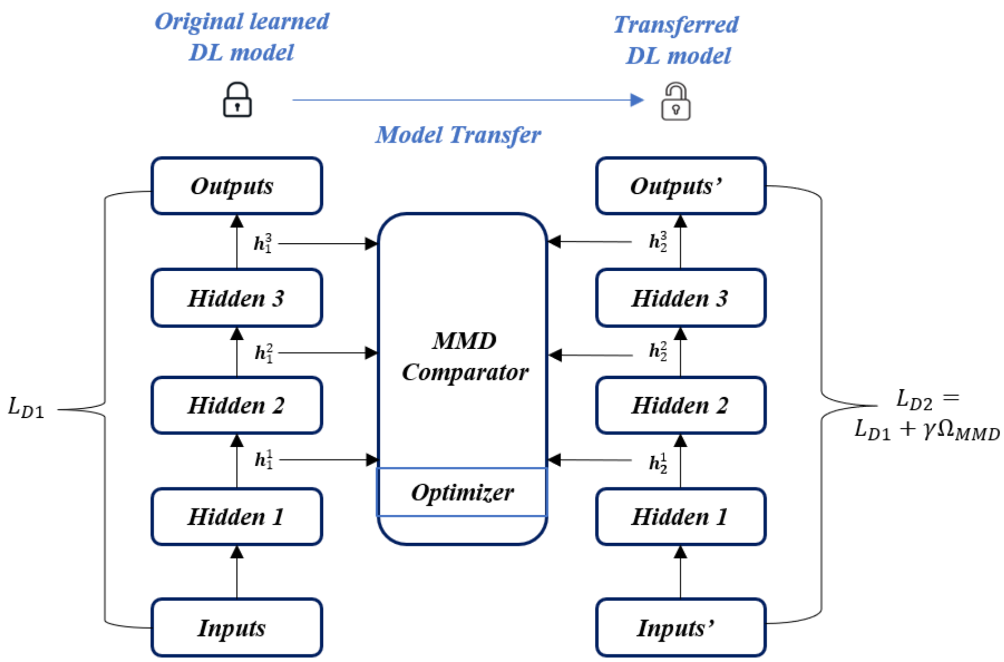


Fig. 1. The architecture of proposed TL with adaptive sparse deep learning

2.2. Knowledge Distillation

Distillation is a chemistry term referring to the extraction of components at different boiling temperatures. Typically, big and complex models that have excellent performance and generalization, but with enormous parameters are hard to be deployed, which is not available to small models. Inspired by pedagogy, in a similar problem, whether the knowledge acquired from big models could guide small ones, so that the latter, meanwhile with reduced parameters' number, have improved performance and generalization ability like the former.

KD [38], based on the "teacher-student" learning strategy, is a model compression approach, and its application is emerging in academia and industry due to its conciseness and effectiveness [39]. Specifically, it means relocating the pre-trained knowledge from teacher networks, typically large or ensemble, to student networks that are relatively small or simple, so that the student's learn the know-how from the teacher's network. The KD system generally involves three pivotal components: knowledge, distillation algorithm, and teacher-student architecture [39].

While former KD studies have focused on classification tasks, few studies concentrated on regression missions, such as speech recognition [40] and objects localization [41], etc. There is a scarcity of investigation on KD approaches for time-series tasks., which are fully developed in the present study.

2.3. Proposed KD forecasting framework

In the present study, inspired by [42], [43], and [44], the KD networks architecture shown in Fig. 2, which is implemented to distill the wind park prediction model with big data to turbine prediction. The proposed KD forecasting framework is developed to address three challenges.

1. Large stacked LSTM-based encoder-decoder models are adopted to predict the wind park with big data while small Bidirectional LSTM (Bi-LSTM) models are applied to forecast small data turbines.
2. Cross-dataset knowledge transfer, model to model, in KD is achieved by developing a suitable operator that allows different teacher and student models, based on different inputs, in the KD structure to be compared in training.
3. The designed linear combination of *Loss* I and II in designing student model's loss function *Loss* KD realizes the optimization of the student model not only by training from its own data but also by further learning from the high-performance complex teacher model.

The operation of this KD architecture is as follows: firstly, the wind park big data is imported into the large teacher model, as the pre-trained, to learn the data depth features and generate the predicted park power. Simultaneously, the turbine data are also fed into the student model for preliminary learning and to produce preliminary turbine power calculations. Then the fine-tuning process with KD for the student model begins with the mission of minimizing its loss function *Loss* KD. In regression tasks with KD, the loss function can be generally expressed in (5)

$$L_{\mathrm{reg}} = \frac{1}{n}\sum_{i=1}^{n} min\left(\|\widehat{\mathbf{p}}_S - \mathbf{p}_S\|^2, \|\widehat{\mathbf{p}}_S - \widehat{\mathbf{p}}_T\|^2\right) \tag{5}$$

where $\widehat{\mathbf{p}}_T$ donates output of the pre-trained teacher model, $\widehat{\mathbf{p}}_S$ is outputs of primarily student model, and $\mathbf{p}_S$ is corresponding real values of inputs of the student model. Notably, since these park and turbine predicted power statistics are not identical, operators are necessary to scale them and to compare each other. The mentioned $\mathbf{p}$ values are defined as in Eq. (6) and calculated with the two operators.

$$\widehat{\mathbf{p}}_T = \left\|\frac{P_p - \hat{P}_p}{P_p}\right\|, \widehat{\mathbf{p}}_S = \left\|\frac{P_t - \hat{P}_t}{P_t}\right\|, \mathbf{p}_S = \mathbf{0} \tag{6}$$

Simplifying the function to a linear combination of two loss functions as in

$$L_{\mathrm{reg}} = \frac{1}{n}\sum_{i=1}^{n} \alpha\|\mathbf{p}_S\|^2 + (1-\alpha)\|\widehat{\mathbf{p}}_S - \widehat{\mathbf{p}}_T\|^2 \tag{7}$$

To prevent overfitting and reduce model complexity, the KD will stop when the student model has a lower error than the teacher one. The formula (7) can be further simplified as

$$L_{\mathrm{reg}} = \frac{1}{n}\sum_{i=1}^{n} \alpha\|\mathbf{p}_S\|^2 + (1-\alpha)L_{\mathrm{compare}} \tag{8}$$

$$L_{\mathrm{compare}} = \begin{cases} \|\widehat{\mathbf{p}}_S - \widehat{\mathbf{p}}_T\|^2, & if\ \|\widehat{\mathbf{p}}_S\| > \|\widehat{\mathbf{p}}_T\| \\ \mathbf{0}, & otherwise \end{cases} \tag{9}$$

where P_p and P_t are measured power of park and turbine, $\hat{P}_p$ and $\hat{P}_t$ and their predicted power by the teacher and student models, respectively. $0 < \alpha < 1$ represents the vital index named KD parameter that controls the balance between knowledge flow from teacher and student itself in the learning stage.

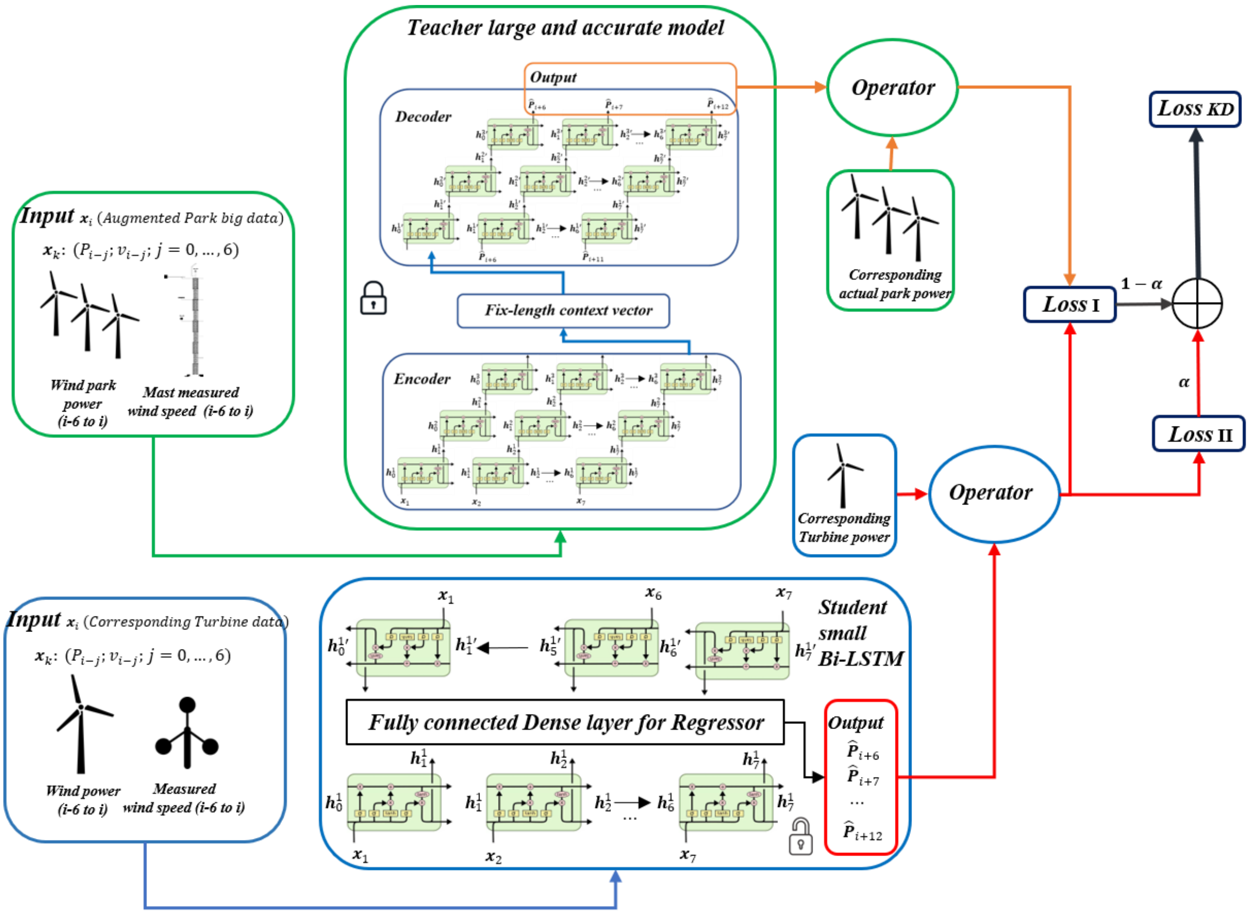


Fig. 2. The proposed KD structure for distilling knowledge from park to turbine.

2.4. Error correction for NWP

As described in Section1, NWP data are often on large scales and can often be downscaled from tens of kilometers to a few ones as applicable to wind parks. However, research has shown that a lower scaling is not always advisable in wind power prediction [45]. Also, it is difficult to accurately downscale NWP to wind turbines (tens of meters) because of the involved turbulence and sophisticated CDF assumptions. In engineering practice, the NWP wind data corresponding to the park are accessible. Meanwhile, the turbine data are analogous to those of the park. In view of the profound pattern mining and matching capabilities of deep learning and the neat structure of TL, park's NWP information can be refined and transferred to more accurate turbine predictions through the corresponding data-driven prediction error correction approach. The proposed NWP error correction approach is explained in detail in section 3.2.

2.5. Modular for Forecasting algorithms

The research employs three deep learning regression algorithms: a multilayer general deep neural network for univariate TL (error correction from park to turbine and integration of turbine training and testing), Bi-LSTM for student prediction of turbines model, and a developed large-scale EDLSTM for forecasting the wind park, respectively. The first two are well-established deep learning models that are well elaborated in [28]. And the superiority of EDLSTM, applied in the proposed KD structure for park prediction algorithms, compared to other machine learning benchmark models (Persistence, simple three-layer backpropagation Neural Networks, basic LSTM, bionic optimized neural networks constructed Adaboost ensemble learning) has been thoroughly analyzed in [17]. And the detailed description of competitors for the proposed prediction framework is described in section 5.2. Since modern deep learning models are modularly constructed, a brief summary of the used modules is presented in Table. 1.

Table 1. A brief summary of prediction modules' setup

Forecasting model	**Main parameters**
Fully connected layer	Sigmoid activation function, and Mean Square Error (MSE) loss function
Deep neural network for univariate TL	The input, 3 hidden, and output layers are with 7, (15, 10, 15) and 7 neurons, respectively; sigmoid activation function, and MSE loss function. (Hidden layers neurons numbers are roughly found with a grid search with a unit of 5 from 5 to 30.)
LSTM unit	Sigmoid activation function, MSE loss function, and Adam algorithm optimizer.
Bi-LSTM	Fully connected layer for internal outputs and the regressor, Two-layer seven LSTM units with reverse directions A fully connected layer with 7 neurons as the external output layer. Dropout avoids overfitting.
EDLSTM	The encoder and decoder, linked with a fixed-length context layer, are with 3 layers of 7 connected LSTM units. Dropout and early stop to avoid overfitting. Batch size control for fast learning.

3. Data preparation and prediction theory

3.1. Data preparation

Wind turbines can convert the kinetic energy from the wind into electricity. Theoretically, the wind power generation model is expressed in (10):

$$P = \frac{1}{2}\rho\pi R^2 C_p(\lambda, \beta) v^3 \quad (10)$$

where P is the wind turbine power generation (W); ρ is the air density (kg/m^2); R donates the rotor radius; C_P means wind energy utilization efficiency (dependent on the tip speed ratio λ and the blade pitch angle β); v represents the wind speed (m/s). According to (10), the output of a wind turbine is mainly about the wind speed, air density, swept area, and blade conditions, all of which are strongly influenced by the local weather of the turbine.

Given that the present investigation is a follow-up study to the previous article [17], the data from the same wind park, a 54 MW with 18 Vestas V90 3 MW turbines with complex terrain in the Arctic, are maintained for comparability. The measured 2017 year data with a 10-mins temporal resolution of the whole wind park (also with 2.5 km spatial and 1 h temporal resolution NWP data) and 5 turbines with varying terrains (T1 Plateau, T2 Valley, T3 Lakeside, T4 Hilltop, T5 Seaside).

For park and turbine, the raw 10-mins measured data derived from the averaged interpolation approach are utilized since it is required to forecast its hours-ahead power. The details of the used data can be found in Section 2 of [17].

Due to the varying physical variables involved in the mentioned equations, the raw data should initially be scaled before learning with data standardization (adjustment by placing the mean at 0 and the standard deviation at 1) to facilitate the convergence speed of the loss function in learning [46].

3.2. Prediction theory

Mathematically, wind power prediction with deep learning can be simplified to a function construction problem. Where the variables can be linked by suitable function operators. This paper proposes the following functional approach to realize transferring large wind park power forecasting models to small turbine prediction and realize non-explicit NWP downscaling by error correction.

Algorithm 1.

1. Teacher (large) models for wind park p and their error functions.

$$\begin{aligned}
\hat{P}_{i+n_p}^{\mathrm{I}} &= f_p^{\mathrm{I}}\left(P_{i-j_p}; v_{i-j_p}\right),\\
Error_p^{\mathrm{I}} &= P_{i+n_p} - \hat{P}_{i+n_p}^{\mathrm{I}},\\
\hat{P}_{i+n_p}^{\mathrm{II}} &= f_p^{\mathrm{II}}\left(P_{i-j_p}; v_{i-j_p}; u_{i+n_p}\right),\\
Error_p^{\mathrm{II}} &= P_{i+n_p} - \hat{P}_{i+n_p}^{\mathrm{II}},\\
Error_p^{\mathrm{II}} &= g_p\left(Error_p^{\mathrm{I}}\right).
\end{aligned} \tag{11}$$

where i is the current time i=1, 2, ..., 7, and with each i, j=0, 1, ..., 6. $\hat{P}_{i+n}$ is n hours ahead predicted wind power, $n \in [6,24]$, v is the wind speed observed in the turbine, u represents the wind speed calculated from the mesoscale NWP wind model for the site.

Firstly, two large deep learning prediction models $f_p^{\mathrm{I}}(.)$ and $f_p^{\mathrm{II}}(.)$ for wind parks are created with only measured data and measured and NWP data, respectively, and the prediction error sequences are calculated separately. Then the relational function $g_p(.)$ of the two error sequences is established.

2. Student (small) training models for turbine t and error functions transfer.

$$\hat{P}_{i+n_{t_{Train}}}^{\mathrm{I}} = f_t\left(P_{i-j_{t_{Train}}}; v_{i-j_{t_{Train}}}\right) \tag{12}$$

$$\hat{P}_{i+n_{t\ Train}}^{\mathrm{II}} = \hat{P}_{i+n_{t_{Train}}}^{\mathrm{I}} + g_p(P_{i+n_t} - \hat{P}_{i+n_{t_{Train}}}^{\mathrm{I}})$$

Firstly, the learned large wind park prediction model $f_p^{\mathrm{I}}(.)$ in the first stage is distilled to the small turbine model $f_t(.)$ with KD. Then the error relational function $g_p(.)$ for park is transferred to non-explicitly compensate for the lack of NWP information in the original prediction model $\hat{P}_{i+n_{t_{Train}}}^{\mathrm{I}}$ for turbines.

3. Turbine testing models

$$\hat{P}_{i+n_{t_{Train}}}^{\mathrm{I}} = \mathfrak{C}_e(\hat{P}_{i+n_{t\ Train}}^{\mathrm{II}}) \tag{13}$$

$$\hat{P}_{i+n_{t_{Test}}}^{\mathrm{I}} = f_t\left(P_{i-j_{t_{Test}}}; v_{i-j_{t_{Test}}}\right)$$

(14)

$$\hat{P}_{i+n_{t\ Test}}^{\mathrm{II}} = \mathfrak{C}^{-1}{}_e(\hat{P}_{i+n_{t_{Test}}}^{\mathrm{I}}) \tag{15}$$

Since the result labels in the test set should not be involved in testing, so the prediction errors in the test set cannot be engaged in the same way as the training for model correction. Therefore, primitive and inverse function transformation is introduced to link the original and improved model (with NWP information) for testing. Firstly, primitive function $\mathfrak{C}_e(.)$ is learned between the two predictions of the training set. Then the trained turbine model $f_t(.)$ is tested on the test set to get an original prediction $\hat{P}_{i+n_{t_{Test}}}^{\mathrm{I}}$. Finally, the improved test set model, delivering the final forecast $\hat{P}_{i+n_{t\ Test}}^{\mathrm{II}}$, is obtained by the inverse operator $\mathfrak{C}^{-1}{}_e(.)$ on $\hat{P}_{i+n_{t_{Test}}}^{\mathrm{I}}$.

To summarize, the foregoing framework for wind turbine power prediction through KD, TL, and error correction is exhaustively illustrated in Fig. 3.

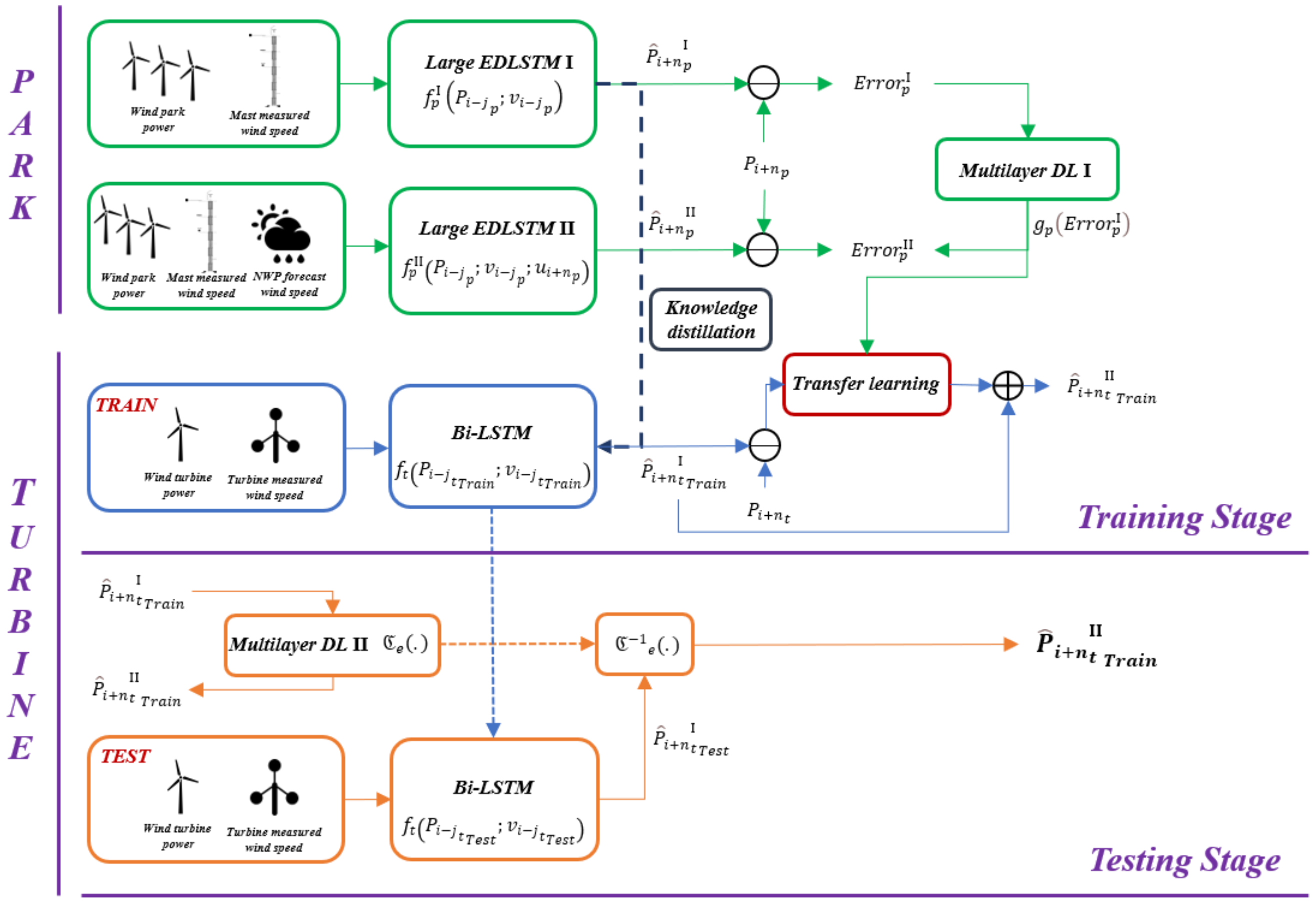


Fig. 3. The proposed KD structure for distilling knowledge from park to turbine.

4. Experiments

4.1. General experimental process

Firstly, the measured wind park and turbine 10-minute data are first interpolated into hourly data by averaging. And all data are standardized and divided into a training set, 65%, and a testing set, 35%. Then add white noise to expand the data to five times the original size (DA2 data augmentation in [17]). The training sets of the park (with 10-folds validations for improving training accuracy for its large forecasting model) and turbines are separately loaded into their corresponding algorithms to learn fine-tuned forecasting models and generate predicted power of the park and turbines. Subsequentially, the prediction errors can be calculated and transferred for the turbine prediction corrections if needed. Finally, the described turbine predictive modeling frames are tested with the turbines' testing sets and the performance is assessed and compared.

4.2. Performance evaluation

The data-driven wind power prediction is essentially categorized as a regression problem in machine learning, where the MSE controls the model's learning progress as the basic unit of the regressor loss function. Therefore, the Root Mean Square Error (RMSE) is the most dominant indicator to evaluate the models' performance. Besides, prediction Qualification Rate (QR) [47] is introduced to more physically evaluate models. RMSE emphasizes larger forecasting errors while the QR stresses smaller ones.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{m}(P_i - \widehat{P_i})^2}{m}} \tag{16}$$

$$QR = \frac{1}{m}\sum_{i=1}^{m}\begin{cases}1, \left(1 - \frac{|P_i - \widehat{P_i}|}{Cap}\right) \geqslant Q \\ 0, \left(1 - \frac{|P_i - \widehat{P_i}|}{Cap}\right) < Q\end{cases} \tag{17}$$

where P_i and $\widehat{P_i}$ represent standardized measured and corresponding predicted wind power, m is the sample number of the testing set. *Cap* donates the designed turbine capacity. Q is the quantile percentage of qualified predictions, selected as 90% in the work as QR90.

Moreover, as a crucial strength of the proposed model is merely training the large EDLSTM network on wind park data and transferring the network's knowledge to the small network for turbine prediction via the developed KD technique, the complexity and computing load of the proposed model could be anticipated to be considerably reduced. So, indices like number of model parameters, modeling time consumption and inference time on the edge are also compared with the rate between forecasting models.

Furthermore, since multiple groups of metrics are involved in the comparisons, a multivariate statistical test, the Friedman test [48], is applied to determine statistical significance in the metrics.

H_o: The column data fail to show a significant difference.

H_a: They display a significant difference.

The statistic F is in (18):

$$F = \frac{12t}{k(k+1)}\left[\sum_{i=1}^{k} r_i^2 - \frac{k(k+1)^2}{4}\right] \tag{18}$$

where t and k represent the number of samples and their columns, r_i is the mean of row i, which follows $\chi^2_{(k-1)}$ distribution under H_o.

5. Results and discussion

This section presents the experimental results from a holistic range in three perspectives. Firstly, given that the KD and TL models serve as core pieces of the proposed method, and the foremost KD parameter α and TL parameter γ remarkably affect the whole forecasting performance, it is necessary to find the most appropriate parameters first. The second part is the main content of the Section, which synthetically evaluates the proposed model by comparing it with the benchmark models. Moreover, a physical analysis of the predicted power is briefly undertaken.

5.1. Optimal TL and KD parameters finding

According to Eq. (7) and (8), the KD parameter α controls the weights of soft-loss from teacher model and hard-loss from student model. Therefore, an optimized KD architecture needs to be constructed before the implementation of the complete turbine prediction modeling. And based on to Eq. (4) the TL parameter γ adjusts the contribution of MMD-based mechanism in TL process, with which the fine-tuned model will be more adaptable to new data.

A grid search is conducted for the parameter in the range of 0 to 1 in steps of 0.2, and the corresponding performance RMSE of the KD turbine prediction model in Fig. 2 (also Algorithm 1.1) with finding α is illustrated in Fig. 4 (a). The overall RMSE, for all targeted turbines, of KD is seen to exhibit nonuniform downward trends as α increases from 0 to 1. However, the RMSE variations and α-values corresponding to the minimum points vary for different turbines. The KD models for the turbines, except for T3 (0.6) and T5 (1), are optimized for an α of 0.8. Intuitively, an interpretation could be that the student model learns more from the complex teacher when the α parameter is bigger, but completely abandoning the student's own loss function may also lead to underfitting on the training set. Therefore, α = 0.8 is chosen as the optimized parameter of KD for all the five turbines in the following trials for γ determinations.

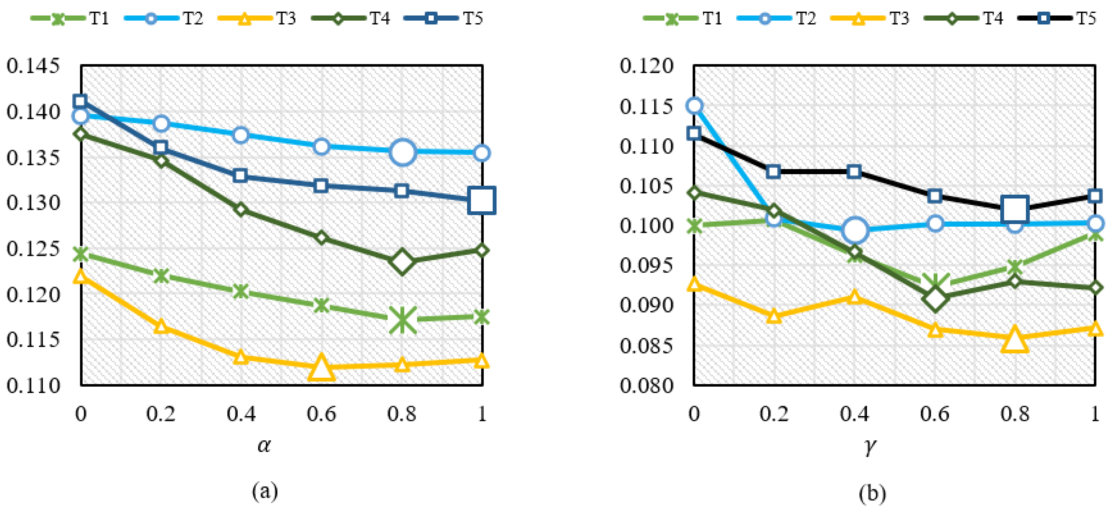


Fig. 4. The grid search for KD and TL parameters with indexing by RMSE (The points corresponding to the best performance is marked with bigger icons): (a) KD parameter α and TL parameter γ, (b) TL parameter γ.

Set the α of KD to 0.8 for all wind turbine prediction frames and then proceed to correct prediction errors with TL from the park to turbines. The same grid search is also made for determining a proper TL parameter γ for error transferring. The performance comparison is displayed in Fig. 4 (b). The RMSE also shows a decrease with increasing γ parameter, but not significantly, and even displays a decreasing and then increasing shape in T1 and T4. Besides, algorithmically, γ controls the contribution of the comparison between wind field and turbine data in the TL, so its optimal value changes with the turbine data and TL internal states. Thus, it is hardly applicable to determine each turbine's TL with a single γ parameter. For T1 and T4, γ is 0.6; for T3 and T5, γ is 0.8; γ is 0.4 for T2.

Therefore, the two dominant parameters for the proposed turbine prediction framework are determined with the grid search and will be applied in the following experiments.

5.2. Multistep turbine power prediction

Two crucial parts are involved in the proposed model, i.e., KD and TL error correction (if the latter is involved, the Algorithm 1.3. must be included for the testing set) for turbine power predictions.

To explore the contribution of each part, several model variants are derived for the comprehensive ablation investigation. The models' description and their inputs are listed in Table 2.

Table 2. The general description for the used models.

Model	**Description**	**Inputs (standardized data)**
BiLSTM	Simple Bi-LSTM for turbine prediction.	Turbine measured data, NWP for park
EDLSTM	[17] proposed EDLSTM framework for turbine prediction.	Turbine measured data, NWP for park
KD	Only Fig. 3 KD system without TL error correction	Park and Turbine measured data
EDED-TL	EDLSTM for both park and turbine forecasting and they are connected with TL error correction, no KD is involved.	Park and Turbine measured data, NWP for park
EDBi-TL	Without KD, EDLSTM for park forecasting, and Bi-LSTM for turbine and they are linked with TL error correction.	Park and Turbine measured data, NWP for park

KD-TL	The proposed predictive framework in which KD combines EDLSTM for park and Bi-LSTM for turbine, their relationship is found in forecasting by TL error correction.	Park and Turbine measured data, NWP for park

The six models and corresponding inputs from the above table are individually applied to five selected wind turbines to predict power from 6 to 12 hours in advance. All prediction models yielded relatively satisfactory outputs within forecasting steps, all with RMSE < 0.15, which fully illustrates the powerful nonlinear mapping capability of the deep LSTM neural network-based models. The RMSE for different models with varying prediction steps is illustrated in Fig. 5.

Generally, all prediction models RMSE rises with increasing steps, but the rising rate is gradually slowing down. In virtually multiple steps, EDED-TL and proposed KD-TL perform best among all turbines' predictions.

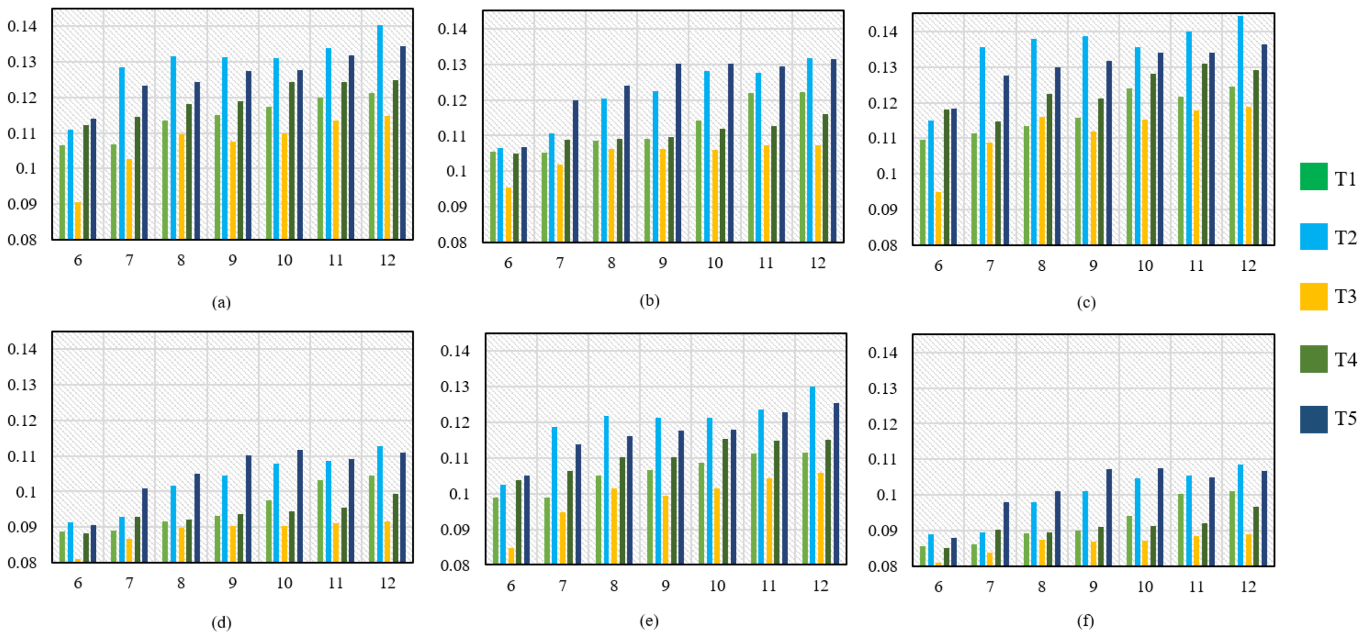


Fig. 5. The multistep RMSE of models with varying prediction algorithms for each turbine: (a) BiLSTM, (b) EDLSTM, (c) KD, (d) EDED-TL, (e) EDBi-TL, (f) KD-TL.

To quantitatively analyze the variations in multistep performance, the RMSE inter-step rising rates are calculated and the summarized geometric means, also with the Friedman tests, and STDs of these rates are taken separately for both turbines and algorithms are shown in the Table 3. Table 3(a) says the p-value of the Friedman statistic is much less than 0.05, indicating that the RMSE increase (from around 1.8% to 3.8%) with the step of the six prediction algorithms is significantly different. This also verifies the dominant influence of topography on turbine predictions. As can be observed from Table 3(b), its Friedman statistic gives a *p*-value of 0.9047 which means the RMSE inter-step increase is statistically identical for the models, which, to some extent, reveals that LSTM-based models can deeply and robustly mine the essential characteristics of the wind data and when a model has an initial edge, the edge carries over.

Table 3(a). The statistics for RMSE inter-step rising rates geomeans for different turbines.

Rising rate between turbines $F = 16.4$; $p = 0.0025$	**T1**	**T2**	**T3**	**T4**	**T5**
GeoMean	2.395%	3.839%	2.749%	1.818%	3.109%
STD	0.0030	0.0032	0.0105	0.0020	0.0043

Table 3(b). The statistics for RMSE inter-step rising rates geomeans for various models.

Rising rate with steps $F = 1.57$; $p = 0.9047$	**BiLSTM**	**EDLSTM**	**KD**	**EDED-TL**	**EDBi-TL**	**KD-TL**
GeoMean	2.853%	2.576%	2.676%	2.716%	2.810%	2.567%
STD	0.0096	0.0082	0.0103	0.0068	0.0097	0.0070

5.3. Holistic validity of the proposed model

Furthermore, the collective performance of various algorithms and different turbines are summarily compared in Fig. 6 with geometric means (including STD) in terms of algorithm and turbine perspectives. Fig. 6 (a) clearly presents the average performance for diverse algorithms in the multistep predictions. KD, despite distilling park's knowledge to turbine, performs the worst, which is because the weather forecasts are not included in the model's inputs, thus the prediction accuracy on historical measured data only is unsatisfactory even with advanced methods. Two baselines with NWP inputs, BiLSTM and EDLSTM, the latter is obviously better than the former, are slightly inferior to EDBi-TL. This indicates that turbine prediction is improved by TL of error correction even without the KD mechanism. And the improvement is considerably bigger than directly incorporating the wind park's NWP data into the turbine predictions. By discovering the advantage of EDED-TL over EDBi-TL, it means that the simple Bi-LSTM, compared to EDLSTM, may not learn the deep features of the turbine data well. Noteworthy, the proposed KD-TL delivers the lowest RMSE, (averagely 21.14%, 7.89%, 23.92%, 3.26%, and 14.86% lower than BiLSTM, EDLSTM, KD, EDED-TL, and EDBi-TL.) which proves the superiority of the proposed models among its competitors.

Fig. 6 (b) summarizes the prediction performance of the algorithms for varying turbines. The RMSE for turbines in flat terrains, such as T1 Plateau and T3 Lakeside, is minimal within an STD. By contrast, the distinctive fjord landscape of the Norwegian coast makes T2 Valley, T4 Hilltop, and T5 Seaside somewhat challenging. But the proposed framework holistically addresses these challenges and provides competitive predictions in these complex terrains (the KD-TL RMSE of these three turbines is smaller than RMSE of some baselines for turbines on flat ground.).

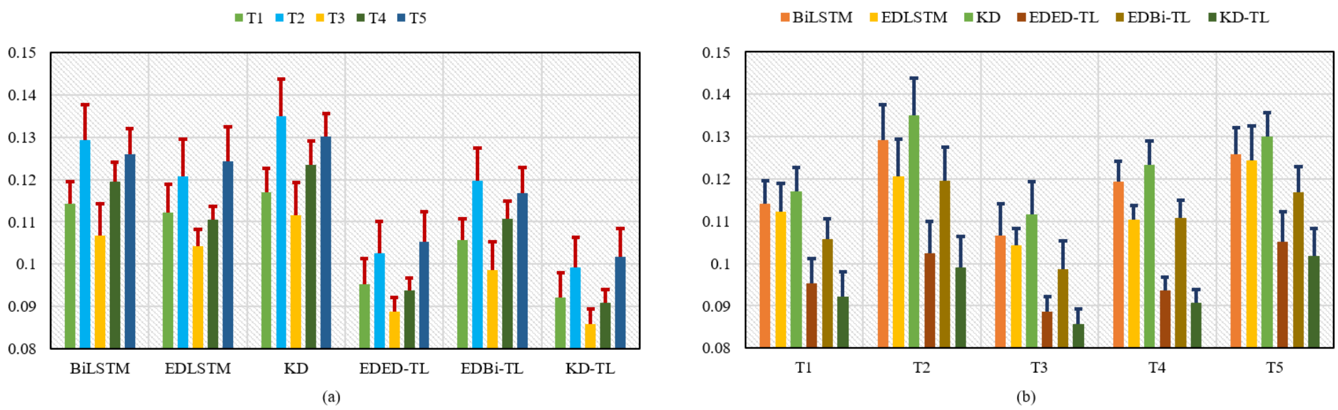


Fig. 6. The overall average RMSE of multistep prediction models with STD: (a) Predictive algorithm perspective, (b) Different turbine perspective.

Finally, the proposed KD-TL has the best performance, even lower than EDED-TL does, both in mentioned two perspectives, which demonstrates that the learned turbine Bi-LSTM (student) with KD from park EDLSTM is even slightly better than the complex EDLSTM, consuming much computing resources and time, for turbine itself. This suggests certain overfitting in the large EDLSTM on the turbine smaller, compared with park, data. Moreover, its advantage over the EDED-TL also exists in model's complexity. The number of EDED-TL parameters is around 3.8 times bigger than the number of KD-TL (also means around 20 times for five turbines' forecasts.). And the former also spends around 2.3 folds as much time as the latter does on edge inference, which meets excepted diminished complexity and rapidity for the proposed model in engineering operations.

To further delve into the turbine power predictions by different. This subsection briefly explores the physical interpretation of the outcomes. Fig. 7 exhibits the geometric QR90 average of five turbines, from 6 to 12 steps, an increase of proposed KD-TL in comparison to other prediction models. As this metric focuses more on predictions within ten percent of the designed power difference between the predicted and real values, a high QR90 reflects that the prediction method gives a more trustworthy result. It is seen that compared to BiLSTM, EDLSTM, KD, EDED-TL, and EDBi-TL, KD-TL offers improvements of QR90 over 14.07%, 11.35%, 14.96%, 1.07%, and 7.86% within one STD respectively.

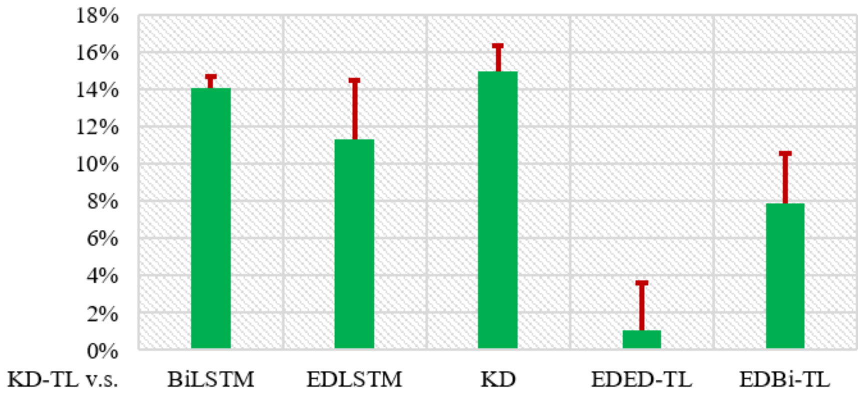


Fig. 7. The overall average QR90 increase rate of the proposed multistep KD-TL prediction versus other models.

In summary, the proposed model enables more precise and economic multistep turbine power prediction than the baseline models and such advantages exist for turbines on different topographic conditions.

6. Conclusions

This paper exploits a deep learning wind power prediction framework. It bridges large (park with big data) and small-scale (turbine with small data) forecasting through a proposed knowledge distillation regression approach, and maps park-scale weather forecasts non-explicitly to turbine-scale by TL-based prediction error corrections. The following conclusions are drawn from experiments on turbine predictions under five types of topography in a park in the Arctic.

As the turbine prediction models get knowledge by KD from the park model, it is vital to optimizing both contributions in the final prediction model (loss function in deep learning). The contributions' parameters for the KD are found as 0.8. And the parameters for error corrections change with turbines and their scope is determined to be 0.4 to 1.

The proposed KD-TL multistep power prediction framework is extraordinarily effective and robust by leveraging the data and reinforcing the nonlinear capabilities of the model based on the experimental comparisons. Compared to its competitors, the overall average effectiveness, by RMSE, is respectively improved from high to low: 23.9%, 21.1%, 14.9%, 7.9%, and 3.3%. The effectiveness is also verified with a metric with physical meaning, QR90. Moreover, KD-TL, thanks to its adequate utilization of weather forecast information, yields satisfactory outcomes in predicting wind turbines in complex terrain, normally challenging, as well. Finally, the complexity and response time of KD-TL decreases multiplicatively over its closest competitor, which enables the proposed approach to have extensive strengths in engineering deployment.

Furthermore, consequent policy recommendations could be inferred:

The assembled historical data should be sufficiently exploited to build accurate large models applicable to the local natural conditions in regional energy industries. With know, a smaller, easy-to-deploy model could be developed in practical applications by edge computing methods. In data-oriented engineering, initial model errors, usually ignored or dismissed as unfavorable, are valuable that can be extracted by progressively state-of-the-art deep learning. The extracted information may help the subsequent modeling through appropriate transfer methodology.

Credit author statement

Hao Chen: Conceptualization, Methodology, Formal analysis, Writing - Original Draft, Writing - Review & Editing; Yngve Birkelund: Comments and Supervision.

References

[1] S. Ahmed, *Wind energy: theory and practice*. PHI Learning Pvt. Ltd., 2015.

[2] W. C. Radünz *et al.*, "The variability of wind resources in complex terrain and its relationship with atmospheric stability," *Energy Conversion and Management,* vol. 222, p. 113249, 2020.

[3] S. K. Salman and A. L. Teo, "Windmill modeling consideration and factors influencing the stability of a grid-connected wind power-based embedded generator," *IEEE Transactions on Power Systems,* vol. 18, no. 2, pp. 793-802, 2003.

[4] H. Liu and C. Chen, "Data processing strategies in wind energy forecasting models and applications: A comprehensive review," *Applied Energy,* vol. 249, pp. 392-408, 2019.

[5] H. Liu, C. Chen, X. Lv, X. Wu, and M. Liu, "Deterministic wind energy forecasting: A review of intelligent predictors and auxiliary methods," *Energy Conversion and Management,* vol. 195, pp. 328-345, 2019.

[6] Z. Qian, Y. Pei, H. Zareipour, and N. Chen, "A review and discussion of decomposition-based hybrid models for wind energy forecasting applications," *Applied energy,* vol. 235, pp. 939-953, 2019.

[7] S. Al-Yahyai, Y. Charabi, and A. Gastli, "Review of the use of numerical weather prediction (NWP) models for wind energy assessment," *Renewable and Sustainable Energy Reviews,* vol. 14, no. 9, pp. 3192-3198, 2010.

[8] J. B. Olson *et al.*, "Improving wind energy forecasting through numerical weather prediction model development," *Bulletin of the American Meteorological Society,* vol. 100, no. 11, pp. 2201-2220, 2019.

[9] J. González-Sopeña, V. Pakrashi, and B. Ghosh, "An overview of performance evaluation metrics for short-term statistical wind power forecasting," *Renewable and Sustainable Energy Reviews,* vol. 138, p. 110515, 2021.
[10] A. Dupré, P. Drobinski, B. Alonzo, J. Badosa, C. Briard, and R. Plougonven, "Sub-hourly forecasting of wind speed and wind energy," *Renewable Energy,* vol. 145, pp. 2373-2379, 2020.
[11] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, "A review of deep learning for renewable energy forecasting," *Energy Conversion and Management,* vol. 198, p. 111799, 2019.
[12] X. Deng, H. Shao, C. Hu, D. Jiang, and Y. Jiang, "Wind power forecasting methods based on deep learning: A survey," *Computer Modeling in Engineering and Sciences,* vol. 122, no. 1, p. 273, 2020.
[13] Y. Wang, R. Zou, F. Liu, L. Zhang, and Q. Liu, "A review of wind speed and wind power forecasting with deep neural networks," *Applied Energy,* vol. 304, p. 117766, 2021.
[14] G. Giebel and G. Kariniotakis, "Wind power forecasting—A review of the state of the art," *Renewable energy forecasting,* pp. 59-109, 2017.
[15] L. Lazić, G. Pejanović, M. Živković, and L. Ilić, "Improved wind forecasts for wind power generation using the Eta model and MOS (Model Output Statistics) method," *Energy,* vol. 73, pp. 567-574, 2014.
[16] W. Y. Cheng, Y. Liu, A. J. Bourgeois, Y. Wu, and S. E. Haupt, "Short-term wind forecast of a data assimilation/weather forecasting system with wind turbine anemometer measurement assimilation," *Renewable Energy,* vol. 107, pp. 340-351, 2017.
[17] H. Chen, Y. Birkelund, and Q. Zhang, "Data-augmented sequential deep learning for wind power forecasting," *Energy Conversion and Management,* vol. 248, p. 114790, 2021.
[18] N. Chen, Z. Qian, I. T. Nabney, and X. Meng, "Wind power forecasts using Gaussian processes and numerical weather prediction," *IEEE Transactions on Power Systems,* vol. 29, no. 2, pp. 656-665, 2013.
[19] Q. Yang, Y. Zhang, W. Dai, and S. J. Pan, *Transfer learning*. Cambridge University Press, 2020.
[20] Q. Hu, R. Zhang, and Y. Zhou, "Transfer learning for short-term wind speed prediction with deep neural networks," *Renewable Energy,* vol. 85, pp. 83-95, 2016.
[21] A. S. Qureshi, A. Khan, A. Zameer, and A. Usman, "Wind power prediction using deep neural network based meta regression and transfer learning," *Applied Soft Computing,* vol. 58, pp. 742-755, 2017.
[22] H. Yin, Z. Ou, J. Fu, Y. Cai, S. Chen, and A. Meng, "A novel transfer learning approach for wind power prediction based on a serio-parallel deep learning architecture," *Energy,* vol. 234, p. 121271, 2021.
[23] X. Liu, Z. Cao, and Z. Zhang, "Short-term predictions of multiple wind turbine power outputs based on deep neural networks with transfer learning," *Energy,* vol. 217, p. 119356, 2021.

[24] B.-M. Hodge *et al.*, "Wind power forecasting error distributions: An international comparison," National Renewable Energy Lab.(NREL), Golden, CO (United States), 2012.
[25] M. Ding, H. Zhou, H. Xie, M. Wu, Y. Nakanishi, and R. Yokoyama, "A gated recurrent unit neural networks based wind speed error correction model for short-term wind power forecasting," *Neurocomputing,* vol. 365, pp. 54-61, 2019.
[26] Z. Sun and M. Zhao, "Short-term wind power forecasting based on VMD decomposition, ConvLSTM networks and error analysis," *IEEE Access,* vol. 8, pp. 134422-134434, 2020.
[27] Z. Liang, J. Liang, C. Wang, X. Dong, and X. Miao, "Short-term wind power combined forecasting based on error forecast correction," *Energy Conversion and Management,* vol. 119, pp. 215-226, 2016.
[28] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
[29] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531,* vol. 2, no. 7, 2015.
[30] Y. Wang, Y. Liu, L. Li, D. Infield, and S. Han, "Short-term wind power forecasting based on clustering pre-calculated CFD method," *Energies,* vol. 11, no. 4, p. 854, 2018.
[31] L. Liu and Y. Liang, "Wind power forecast optimization by integration of CFD and Kalman filtering," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects,* vol. 43, no. 15, pp. 1880-1896, 2021.
[32] X. Ying, "An overview of overfitting and its solutions," in *Journal of Physics: Conference Series*, 2019, vol. 1168, no. 2: IOP Publishing, p. 022022.
[33] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*. Springer, 2013.
[34] C. Ren and Y. Xu, "Transfer learning-based power system online dynamic security assessment: Using one model to assess many unlearned faults," *IEEE Transactions on Power Systems,* vol. 35, no. 1, pp. 821-824, 2019.
[35] L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*: IGI global, 2010, pp. 242-264.
[36] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *International conference on artificial neural networks*, 2018: Springer, pp. 270-279.
[37] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Transactions on systems, man, and cybernetics: systems,* vol. 49, no. 1, pp. 136-144, 2017.
[38] C. Buciluǎ, R. Caruana, and A. Niculescu-Mizil, "Model compression," in *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006, pp. 535-541.
[39] J. Gou, B. Yu, S. J. Maybank, and D. Tao, "Knowledge distillation: A survey," *International Journal of Computer Vision,* vol. 129, no. 6, pp. 1789-1819, 2021.

[40] L. Yuan, F. E. Tay, G. Li, T. Wang, and J. Feng, "Revisiting knowledge distillation via label smoothing regularization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3903-3911.
[41] G. Chen, W. Choi, X. Yu, T. Han, and M. Chandraker, "Learning efficient object detection models with knowledge distillation," *Advances in neural information processing systems,* vol. 30, 2017.
[42] M. R. U. Saputra, P. P. De Gusmao, Y. Almalioglu, A. Markham, and N. Trigoni, "Distilling knowledge from a deep pose regressor network," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 263-272.
[43] Q. Xu, Z. Chen, M. Ragab, C. Wang, M. Wu, and X. Li, "Contrastive adversarial knowledge distillation for deep model compression in time-series regression tasks," *Neurocomputing,* 2021.
[44] M. Kang and S. Kang, "Data-free knowledge distillation in neural networks for regression," *Expert Systems with Applications,* vol. 175, p. 114813, 2021.
[45] J. B. Bremnes and G. Giebel, "Do regional weather models contribute to better wind power forecasts," *The Norwegian Meteorological Institute,* 2017.
[46] H. Chen and R. Staupe-Delgado, "Exploiting more robust and efficacious deep learning techniques for modeling wind power with speed," *Energy Reports,* vol. 8, pp. 864-870, 2022.
[47] L. Lijuan, L. Hongliang, W. Jun, and B. Hai, "A novel model for wind power forecasting based on Markov residual correction," in *IREC2015 The Sixth International Renewable Energy Congress*, 2015: IEEE, pp. 1-5.
[48] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference: Revised and Expanded.* CRC press, 2014.