Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Data science in wind energy: a case study for Norwegian offshore wind

To cite this article: Hao Chen *et al* 2023 *J. Phys.: Conf. Ser.* **2638** 012013

Data science in wind energy: a case study for Norwegian offshore wind

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

[1] Department of Technology and Safety

[2] Department of Physics and Technology

[3] Department of Computer Science

[1,2,3] Faculty of Science and Technology, UiT The Arctic University of Norway, Tromsø 9019, Norway

hao.chen@uit.no

Abstract: In the digital and green transitions, rapidly growing renewable energies are accumulating more and more data. Big data gives room to apply emerging data science to solve challenges in the energy sector. Offshore wind power receives accelerating attention due to its sufficient resources and cleanness. This paper uses data science, including statistical analysis and machine learning, to systematically analyse three coastal wind sites in Norway. The results show that although Norway possesses ample offshore resources, its development could be improved by natural, technical, and economic challenges that can be addressed with the help of data science. Technically, the statistical attributes and forecasting intricacy of offshore wind resources differ across various regions of Norway.

1. Introduction

Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report delivered an updated global assessment of climate change mitigation progress and pledges [1]. The twin transitions of green and digital are two of the most critical challenges and opportunities for human society in the coming years [2]. A concept emerging at the nexus of these two transitions is the so-called *twin transitions.* In energy sectors, the digital transition can help energy utilization be smarter, greener, more sustainable, and more resilient with the advances of data science and multisource information integration into the energy physical system. More specifically, in one of the primary sources of renewable energy - wind energy, data science has become an increasingly crucial tool in analyzing wind energy operations and wind resource assessment [3].

The success of wind energy projects relies heavily on accurate wind energy-analysis, which requires high-quality data. Before building a wind park, the developer must study the site's wind resources to determine the potential for wind power generation. This process involves collecting wind speed and direction data over a while, typically a year or more. The data are then analyzed to determine the average wind speed, wind direction frequencies, and other parameters critical to the wind turbine performance [4].

Offshore wind power presents a promising source of renewable energy in Norway [5]. However, the harsh weather conditions in the Nordic, including extreme temperatures and high winds, can damage turbines and make maintenance challenging. The remote location of offshore wind turbines in the

Norwegian fjords makes it challenging to operate and maintain offshore wind projects quickly, increasing downtime and reducing overall project efficiency. Finally, the cost of offshore wind power remains relatively high compared to other renewable energy sources [6], making it less competitive in the energy market. To maximize the potential of offshore wind in Norway, the emerging data science techniques can provide valuable insights into wind power project design, operational performance and enable proactive maintenance strategies.

Data science is an interdisciplinary field that combines statistical analysis, machine learning, and computer science techniques to extract patterns and insights from complex data sets [7]. In offshore wind power, data science can analyze data collected from weather models, sensors, and other sources to optimize wind turbine and power design, predict power generation and maintenance needs, and improve offshore wind system efficiency.

This paper systematically uses data science to analyze Norwegian offshore wind power, primarily focusing on wind resource assessment and power forecasting.

2. Norwegian Offshore Wind and Related Data Science

2.1. Offshore Wind Energy in Norway

Norway has vast natural sea resources and great offshore and onshore wind energy potential [8]. The country has a long 83, 000 km coastline and exposure to both the North Sea and the Norwegian Sea and complex mountainous terrain. Wind energy has become an essential part of Norway's energy generation mix, accounting for over 10% of the country's total electricity production in 2020. The average wind speeds along the coast vary from 5-10 m/s which is interesting for offshore wind turbines. According to the Norwegian Energy Directorate (NVE), Norway's potential capacity for offshore wind power generation is estimated at 45 GW. This estimate is based on the technical potential of offshore wind, which considers factors such as wind speed, water depth, and distance from shore.

Currently, Norway has one operational offshore wind park (a 2.3 MW Hywind Scotland project located off the coast of Scotland) [9, 10]. However, several projects are in the planning phase, like the Sørlige Nordsjø II project with a planned capacity of 1.5 GW and the 88 MW Havsul 1 project. The Norwegian government has set a target of 4-6 TWh of electricity generated from offshore wind by 2025 and aims to become a leading exporter of offshore wind power technology [11]. Several projects are in the planning phase, and with the proper regulatory framework and investment, offshore wind power could play a significant role in achieving Norway's energy goals.

2.2. Data Science in Wind Energy

With the growing data from meteorological stations, remote sensing, wind turbines, and other measurements, there are huge potentials with data and information sciences to improve our understanding of wind resources and patterns and optimize wind energy utilization design and operation. Emerging data science, including advanced statistical analysis and machine learning, has become increasingly important in analyzing wind energy.

One of the critical applications of data science in wind energy is wind resource assessment. Statistical and machine learning methods are widely used in the topic. Analyzing data from various sources, including meteorological stations, remote sensing instruments such as lidar or radar, and numerical weather prediction models, can create a more comprehensive picture with statistics of the wind conditions at a particular site. It allows us to make more accurate predictions about potential wind energy generation and optimize wind turbines' design and placement [12,13]. Machine learning algorithms are also particularly useful, as they can analyze large amounts of data and identify hidden and complex patterns that may not be apparent using traditional statistical methods [14]. Deep learning can identify patterns in historical wind data and predict future wind conditions. At the same time, neural networks can classify different wind patterns, such as gusts or turbulence [15]. Another essential data science application in wind energy is power forecasting. Wind power forecasts are vital for ensuring a stable and reliable power supply from wind turbines, as they allow grid operators to prepare for fluctuations in wind power output. Finally, the forecasts can help to optimize wind park operations and improve the overall efficiency of energy systems by analyzing historical wind and meteorological data with machine learning algorithms [16].

3. Applied Data Science Methodologies

This section briefly describes the statistical and machine learning algorithms applied in the following case study.

3.1. Statistical Methods

This paper employs descriptive statistics analysis, probabilistic modeling, and hypothesis testing to analyze offshore wind.

Descriptive statistics are widely used in statistical analysis and can be found in [17]. The Probability Density Function (PDF) describes the likelihood of a random variable occurring at a specific point within each observation interval. This work uses the generalized extreme value distribution to model the PDF of the wind speed as:

$$f(x \mid k, \mu, \sigma) = \left(\frac{1}{\sigma}\right) \exp\left(-\left(1 + k\frac{(x-\mu)}{\sigma}\right)^{-\frac{1}{k}}\right)\left(1 + k\frac{(x-\mu)}{\sigma}\right)^{-1-\frac{1}{k}} \tag{1}$$

where k is a shape parameter, μ is a location parameter, and $\sigma$is a scale parameter.

Two important hypothesis tests (Augmented Dickey-Fuller test (ADF) [18] and Kolmogorov-Smirnov test [19] (K-S test) [19]) are also employed to check the offshore wind data.

Stationarity Test: The Augmented Dickey-Fuller test (ADF) tests.

H_0: The data are non-stationary. H_a: The data are stationary.

Goodness-of-fit Test: The two-sample Kolmogorov-Smirnov test (K-S test).

H_0: Two datasets have the same distribution. H_a: They do not have the same distribution.

3.2. Machine Leaning Methods

Machine learning is an algorithm that allows computers to learn patterns and insights from data and make predictions or decisions based on the learning. Machine learning is widely used in various fields, such as image and speech recognition, natural language processing, fraud detection, and many others [20] Machine learning algorithms for regression use statistical models to identify the relationship between the input variables and the output variable in cases of continuous numeric values and then use this model to predict the value of the output variable with new input values [21].

The forecasting of future wind speed is based on autoregression in the study, such as in Equation (2):

$$X_{t+n} = f\left(X_{t,}X_{t-1,}X_{t-2,} \cdots, X_{t-p}\right) + \varepsilon \tag{2}$$

The autoregression can be recognized with machine learning algorithms. This investigation uses four representative machine learning, including linear and different neural networks, to forecast offshore wind speed. Linear Regression (LR), back propagation Neural Network (NN) with 12 hidden neurons, CNN Convolutional Neural Networks (CNN), and LSTM Long Short-Term Memory (LSTM) are quite mature machine learning algorithms, and their introduction can be found in [22] due to space reasons.

4. Case Study

This paper selects three sites along the Norwegian coastline for case studies on data science in offshore wind energy. Namely, Torsvåg Fyr (70.5882N, 19.2844E) in Northern Norway, Myken (66.7621N, 12.4809E) in Middle Norway, and Utsira Fyr (59.3102N, 4.8857E) in Southern Norway. The wind speed data are extracted from the Norsk klimaservicesenter; they range from January 1, 2020, to December 31, 2022. The raw wind speed data are shown in Figure 1. Table 1 gives the descriptive statistics for these wind speed data.

Figure 1. Wind speed of three sites on Norway coast (Norsk klimaservicesenter)

Table 1. Descriptive statistics of wind speed of three sites on Norway coast

Statistics / Wind sites	Mean (m/s)	Median (m/s)	Standard Deviation (m/s)	Max (m/s)	Min (m/s)	Skewness	Kurtosis
Torsvåg Fyr	6.0436	5.5	3.875	28.5	0	0.825	3.5492
Myken	8.3283	7.8	4.1467	27.1	0.2	0.6163	3.2163
Utsira Fyr	8.3948	7.8	4.3910	28.9	0.1	0.6549	3.0848

It can be seen that the Torsvåg Fyr site in the North has a lower average wind speed, but it also has a maximum wind speed of nearly 30 meters per second, indicating that the overall wind resources of the site are not as good as the two sites to its south. This site has a large Skewness and Kurtosis, suggesting that its wind speed distribution is relatively very concentrated in the low wind speed region and has a longer high-speed tail. Myken site in the Middle and Utsira Fyr in the south have a similar historical distribution of wind speeds but are more stationary compared to the Northern site. The estimated pdf for each site is shown in Figure 1 with the scaled histogram resulting from the measured wind speed, and Table 2 summarizes the resulting parameters. The pdf of Torsvåg Fyr does match the high occurrence of very low wind speeds in the range of 1-2 m/s (or the overestimation for 3-5 m/s), but it gives better results for wind speeds higher than 6 m/s. It can also be observed that similar wind speeds at Myken and Utsira Fyr.

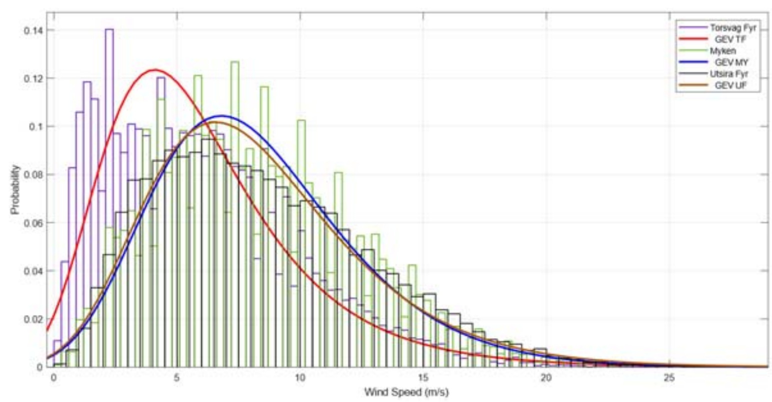

Figure 2. Generalized extreme value distribution modeling of wind speed of the three sites

Table 2. Generalized extreme value distribution modeling parameters

Generalized extreme value distribution	Torsvåg Fyr	Myken	Utsira Fyr
Distribution mean (m/s)	6.0339	8.3103	8.3696
Distribution Standard Deviation (m/s)	4.024	4.146	4.423
Shape parameter	0.0370	-0.0762	-0.0391

Location parameter	4.2000	6.5186	6.4156
Scale parameter	2.9813	3.5360	3.6192

The K-S tests show that all three sites do not follow the same distribution, with all K-S tests *p*-values significantly smaller than 0.05, as shown in Figure 1. Further, to perform the time series analysis of the wind speed, it is necessary to test the series stationarity with the ADF test; the results are shown in Table 3. The *p*-values of the wind speed data for all three offshore wind sites are significantly less than 0.05, indicating that all three-time series are stationary. No annual or quarterly division is required in the analysis and forecasting.

Table 3. ADF test *p*-values and statistics

Wind sites	**Torsvåg Fyr**	**Myken**	**Utsira Fyr**
ADF *p*-values	0.0001	0.0001	0.0001
ADF *Statistics*	-22.0376	-14.1959	-11.9411

Finally, the four machine learning algorithms mentioned above are used to predict three sites' wind speeds from 1 to 3 hours ahead, where the inputs are the wind speeds recorded in the previous six hours from the current. Root Mean Square Error (RMSE) ($\boldsymbol{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (y(t) - \hat{y}(t))^2}$) and correlation coefficient between predicted speeds and the ground-truth speed values are employed to show the forecasting performance, which is displayed in Table 4. As the forecast time lengthens, the wind speed forecast performance decreases significantly. The correlation for T+1, 2, and 3 gives essential information about autocorrelation. Overall, LSTM networks have the best prediction performance. This may be because they greatly capture the time dependence in the wind data so that the algorithms can be selected for further optimization in offshore wind prediction. Despite their complex network structure, CNNs perform much worse in making univariate wind speed predictions than image recognition. It is worth noting that the relatively simple linear regression performs almost as well as the neural network. This may be due to the short forecasting time; wind speed is linear in the period. Notably, NN performs nearly as well as LSTM in onshore wind prediction [23], but LSTM delivers better forecasts in offshore cases. This may reflect the complexity of offshore wind data, which is more time dependent. Further, the Torsvåg Fyr site in the North has worse forecasting performance than the two sites to its south, which shows that it is more challenging to predict the highly volatile Arctic offshore winds.

Table 4. Wind speed forecasting of three sites on the Norway coast

RMSE				**Correlation coefficient**			
Torsvåg Fyr	**T+1**	**T+2**	**T+3**	**Torsvåg Fyr**	**T+1**	**T+2**	**T+3**
LR	1.8505	2.2193	2.4966	LR	0.8812	0.8239	0.7719
NN	1.8460	2.2121	2.4915	NN	0.8818	0.8251	0.7730
CNN	1.8462	2.2176	2.4962	CNN	0.8817	0.8242	0.7720
LSTM	1.7805	2.1518	2.4097	LSTM	0.9005	0.8367	0.7831
Myken	**T+1**	**T+2**	**T+3**	**Myken**	**T+1**	**T+2**	**T+3**
LR	1.5751	1.9930	2.3210	LR	0.9224	0.8745	0.8242
NN	1.5703	1.9735	2.3022	NN	0.9229	0.8771	0.8273
CNN	1.5774	1.9927	2.3186	CNN	0.9222	0.8746	0.8246
LSTM	1.5537	1.9630	2.2578	LSTM	0.9431	0.8917	0.8425
Utsira Fyr	**T+1**	**T+2**	**T+3**	**Utsira Fyr**	**T+1**	**T+2**	**T+3**
LR	1.3350	1.7384	2.1044	LR	0.9542	0.9202	0.8832
NN	1.3313	1.7178	2.0637	NN	0.9545	0.9222	0.8880
CNN	1.3367	1.7329	2.0923	CNN	0.9541	0.9207	0.8847
LSTM	1.3219	1.6910	2.0397	LSTM	0.9724	0.9375	0.8981

5. Conclusion

This paper discusses the Norwegian offshore wind resources and development and the data science applied to wind energy. Statistical and machine learning analyses of wind speed data from three sites in northern, central, and southern Norway are also presented as a case study. The case study shows the potential of offshore wind in some Norwegian coastal areas and how machine learning methods can be used in univariate forecasting of wind speeds. As per the study's results, the following conclusions can be drawn: 1. Norway has good offshore resources, but there are still difficulties in developing them due to natural, technical, and economic reasons, where data science can contribute to tackling these difficulties. 2. Offshore wind resources in different parts of Norway vary in their statistical characteristics and forecasting complexity (the Arctic region is more difficult to predict, and there are more abnormal events). Developing methodologies for coordinating these offshore wind analyses is the next research priority.

Acknowledgments

The author is funded by Equinor's Academia Programme through a collaboration agreement with UiT The Arctic University of Norway. Thanks Dr. Benjamin Ricaud for his insightful comments and thanks Department of Technology and Safety and Department of Physics and Technology, UiT NT-fak, for administration support.

References

[1] A. Ming, I. Rowell, S. Lewin, R. Rouse, T. Aubry, and E. Boland, "Key messages from the IPCC AR6 climate science report," 2021.
[2] R. Lema and R. Rabellotti, "The green and digital transition in manufacturing global value chains in latecomer countries," 2023: UNCTAD United Nations Conference on Trade and Development.
[3] P. Veers et al., "Grand challenges in the science of wind energy," Science, vol. 366, no. 6464, p. eaau2027, 2019.
[4] M. Islam, R. Saidur, and N. Rahim, "Assessment of wind energy potentiality at Kudat and Labuan, Malaysia using Weibull distribution function," Energy, vol. 36, no. 2, pp. 985-992, 2011.
[5] K. Meier, "Hydrogen production with sea water electrolysis using Norwegian offshore wind energy potentials: Techno-economic assessment for an offshore-based hydrogen production approach with state-of-the-art technology," International Journal of Energy and Environmental Engineering, vol. 5, pp. 1-12, 2014.
[6] Z. Ren, A. S. Verma, Y. Li, J. J. Teuwen, and Z. Jiang, "Offshore wind turbine operations and maintenance: A state-of-the-art review," Renewable and Sustainable Energy Reviews, vol. 144, p. 110886, 2021.
[7] A. Blum, J. Hopcroft, and R. Kannan, Foundations of data science. Cambridge University Press, 2020.
[8] R. Pelc and R. M. Fujita, "Renewable energy from the ocean," Marine Policy, vol. 26, no. 6, pp. 471-479, 2002.
[9] A. Jacobsen and M. Godvik, "Influence of wakes and atmospheric stability on the floater responses of the Hywind Scotland wind turbines," Wind Energy, vol. 24, no. 2, pp. 149-161, 2021.
[10] S. Whitfield, "Offshore Wind: The New Frontier in Powering Platforms?," Journal of Petroleum Technology, vol. 72, no. 01, pp. 38-40, 2020.
[11] L. H. Gulbrandsen, T. H. J. Inderberg, and T. Jevnaker, "Is political steering gone with the wind? Administrative power and wind energy licensing practices in Norway," Energy Research & Social Science, vol. 74, p. 101963, 2021.
[12] K. Murthy and O. Rahi, "A comprehensive review of wind resource assessment," Renewable and Sustainable Energy Reviews, vol. 72, pp. 1320-1342, 2017.

[13] 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, p. 7613, 2021.
[14] J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng, "A survey of machine learning for big data processing," EURASIP Journal on Advances in Signal Processing, vol. 2016, pp. 1-16, 2016.
[15] 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.
[16] 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.
[17] T. G. Nick, "Descriptive statistics," Topics in biostatistics, pp. 33-52, 2007.
[18] W. H. Greene, Econometric analysis. Pearson Education India, 2003.
[19] A. Justel, D. Peña, and R. Zamar, "A multivariate Kolmogorov-Smirnov test of goodness of fit," Statistics & Probability Letters, vol. 35, no. 3, pp. 251-259, 1997.
[20] Z.-H. Zhou, Machine learning. Springer Nature, 2021.
[21] S. Rong and Z. Bao-Wen, "The research of regression model in machine learning field," in MATEC Web of Conferences, 2018, vol. 176: EDP Sciences, p. 01033.
[22] G. Bonaccorso, Machine learning algorithms. Packt Publishing Ltd, 2017.
[23] H. Chen and Y. Birkelund, "An Evaluation on Diverse Machine Learning Algorithms for Hourly Univariate Wind Power Prediction in the Arctic," in Journal of Physics: Conference Series, 2021, vol. 2141, no. 1: IOP Publishing, p. 012016.