

Extracting Data-Driven Insights from LiDAR and SCADA Data for Improved Analysis of Offshore Wind Farms

H H Mian¹, B E Rognskog Nordbak¹, E Skoglund¹, T Asim², L Yang³,
M S Siddiqui¹

¹Department of Mechanical Engineering and Technology Management, Norwegian University of Life Sciences, 1432, Ås, Norway

²School of Computing, Engineering, and Technology, Robert Gordon University, Aberdeen, UK

³Division of Energy and Sustainability, School of Water, Energy and Environment (SWEE), Cranfield University, Bedford MK43 0AL, United Kingdom

E-mail: Corresponding Author (H H Mian): haris.hameed.mian@nmbu.no

Abstract. Offshore wind energy is essential for the global transition to renewable energy. Wind farms, such as those at Anholt and Westernmost Rough, are crucial in providing clean electricity. LiDAR (Light Detection and Ranging) and SCADA (Supervisory Control and Data Acquisition) data are particularly valuable for understanding turbine behavior and predicting energy output. Recent advancements in machine learning (ML) and artificial intelligence (AI) have opened new possibilities for analyzing complex wind farm data. This study employs data-driven techniques, specifically XGBoost and Long Short-Term Memory (LSTM) networks, to enhance the analysis of LiDAR and SCADA data from the Anholt and Westernmost Rough offshore wind farms. Data-driven filters were applied to enhance the quality of input data, thereby improving the accuracy of the models and reducing noise. XGBoost demonstrated computational efficiency, training faster than Bi-LSTM while achieving an R^2 of 0.97 for Anholt and 0.86 for Westernmost Rough. While Bi-LSTM successfully captured temporal dependencies, it required significantly longer training times. RMSE and MSE results indicate that XGBoost outperformed Bi-LSTM at Anholt by 6.3% and 12.2%, respectively, whereas both models showed higher errors at Westernmost Rough, likely due to data dependency. The residual analysis confirmed tighter error distribution for Anholt, whereas Westernmost Rough exhibited higher prediction uncertainties. Wind speed loss analysis revealed that turbines in the selected rows experienced variations, highlighting the impact of local turbulence on wind flow characteristics.

Keywords: floating Offshore wind, data-driven, LiDAR data, SCADA data, Bi-LSTM, XGBoost, wind farm

1. Introduction

The transition to renewable energy has become a global priority, driven by the urgent need to combat climate change and reduce dependence on fossil fuels. Offshore wind development is rapidly expanding, offering sustainable energy solutions by harnessing stronger and more consistent wind resources [1]. Wind power prediction is categorized into short-term, medium-term, and long-term predictions based on the time scale analysis. Short-term wind power prediction, typically 30 minutes to 6 hours, is important for optimizing wind farm operations



and predicting energy generation [2]. It relies on different data sources, including meteorological tower measurements of wind speed, direction, temperature, humidity, etc., at wind farm location [3]. While atmospheric stability and turbulence significantly influence wind power generation, their use in prediction models has been limited due to measurement challenges [4]. LiDAR (Light Detection and Ranging) is a remote sensing technology that measures wind conditions [5]. Still, its application in predictive wind power research is limited, especially in challenging weather and terrain. Most prediction studies rely on SCADA (Supervisory Control and Data Acquisition) [6] and meteorological tower data, focusing little on integrating LiDAR for short-term predictions. A gap remains in evaluating LiDAR's accuracy and its potential improvement of prediction models when combined with SCADA data.

Most existing studies rely on short-term data (days to months) and focus on a small number of turbines [7], while research using long-term data (>1 year) for large offshore wind farms remains limited. This gap affects the reliability of wind power plant operation predictions and strategy development. Machine learning models have gained traction in wind power prediction [8], but prior studies have not fully leveraged integrated LiDAR and SCADA data, limiting their ability to reduce uncertainty and enhance accuracy. To address these challenges, we developed short-term wind power prediction models using Bi-LSTM [9, 10] and XGBoost [11] with LiDAR and SCADA data. Bi-LSTM effectively captures temporal dependencies in sequential data, making it well-suited for wind power prediction. Meanwhile, XGBoost efficiently handles structured data and feature interactions, thereby enhancing predictive accuracy. We utilized offshore wind farm datasets from Anholt (ANH) and Westernmost Rough (WMR) to compare model performance and assess generalization across different wind farms. The models were evaluated based on training time, RMSE, MSE, and R^2 , providing insights into their accuracy and computational efficiency. This study uniquely combines SCADA and LiDAR data, addressing limitations of prior models that typically utilize only one data source for short-term wind power forecasting. In addition to power prediction, the study also investigates wind speed loss through a comparative analysis of LiDAR and SCADA measurements. This analysis offers insight into the effects of local turbulence and discrepancies between sensor types, which directly impact feature quality and model performance. The models developed in this study aim to enhance short-term (1-hour ahead) power prediction by incorporating lagged and statistical features from historical SCADA and LiDAR measurements, rather than relying solely on instantaneous wind-speed observations. The paper is structured as follows: Section II presents the wind farms and datasets used in this study, with a focus on the Anholt and Westernmost Rough wind farms. Section III details the methods employed, describing the XGBoost and Bi-LSTM models used for wind power predictions. Section IV discusses the results and analysis, covering data filtering techniques, power prediction performance, and wind speed loss. Finally, the findings are summarized, highlighting the comparative performance of the models and their applicability to offshore wind power prediction.

2. Wind farms and dataset

2.1. Anholt wind farm

The Anholt Offshore Wind Farm [12], situated in the Kattegat Sea between Djursland and the island of Anholt, is one of Denmark's largest offshore wind farms. Operational since 2013, it comprises 111 Siemens Gamesa SWT-3.6-120 turbines, each with a capacity of 3.6 MW, totaling approximately 400 MW. The wind farm layout is shown in Figure 1(a). The farm spans an area of about 88 km², measuring approximately 20 km in length and 5 km in width, with water depths ranging from 15 to 19 meters. It is roughly 15 km from Djursland to the southwest and about 20 km from Anholt Island to the northeast. The project's development was led by Ørsted (formerly DONG Energy), with Siemens contracted to supply the turbines. The collected data from this farm include LiDAR, buoy, and SCADA data; however, only LiDAR and SCADA data

were used in this study.

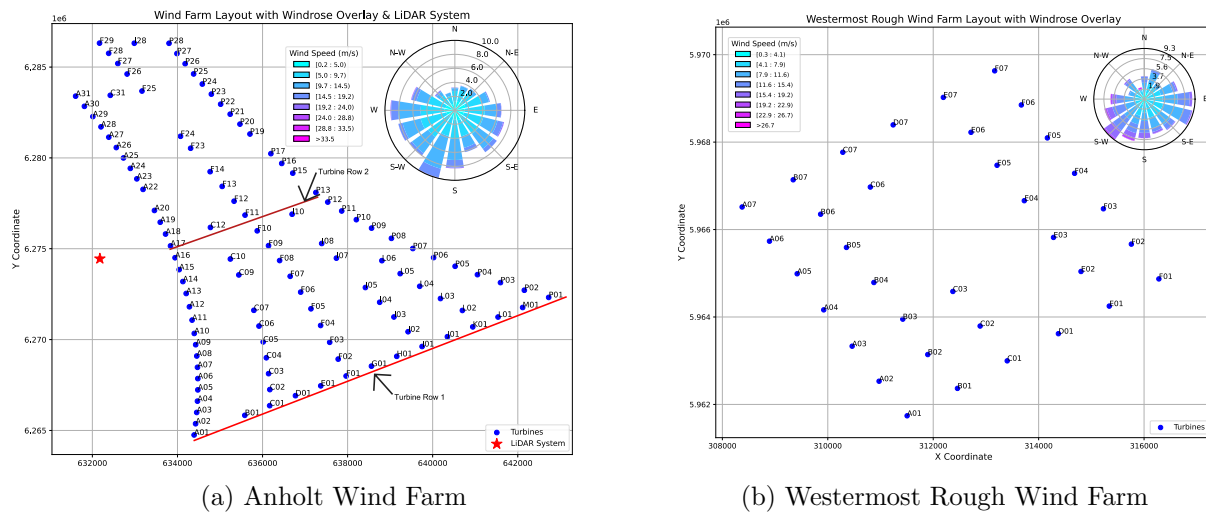


Figure 1: Wind farm layout visualizations for Anholt and Westermost Rough. The wind roses for these sites are also shown in their respective layout. LiDAR systems were installed at both sites.

2.2. Westermost Rough wind farm

The Westermost Rough Offshore Wind Farm is located in the North Sea, approximately 8 km northeast of Withernsea, off the Holderness coast in England. Covering an area of about 35 km², the wind farm has an installed capacity of 210 MW, supplying electricity to over 150,000 households. The wind farm layouts are shown in Figure 1(b). The project was initially awarded to Total in 2003 but was later taken over by DONG Energy (now Ørsted) in 2007. The farm consists of 35 Siemens SWT-6.0-154 wind turbines, each with a capacity of 6 MW and a rotor diameter of 154 meters. These turbines are supported by monopile foundations with diameters of 6.5 meters and weights up to 800 tonnes. Construction began with the installation of the first foundation in February 2014, and the wind farm became fully operational in May 2015. Although not shown explicitly in Figure 1(b), a Leosphere WindCube LiDAR system was deployed at the Westermost Rough site as well. LiDAR data collected at multiple heights were included in this study alongside SCADA data, similar to the Anholt wind farm.

2.3. Dataset description and preprocessing overview

A comparison of key parameters for both wind farms is given in Appendix 'A'. We utilized LiDAR and SCADA datasets provided by Ørsted for both the Anholt and Westermost Rough wind farms. These datasets consist of raw LiDAR measurements, including wind speed, direction, turbulence intensity, and environmental conditions recorded at multiple heights. LiDAR systems from Leosphere WindCube were installed at both sites for wind assessment. The data includes timestamps, external temperature, air pressure, relative humidity, and wind speed parameters such as mean, dispersion, minimum, and maximum values. Additionally, the dataset contains information on vertical wind speed, Doppler spectral broadening, and data availability recorded at 12 different measurement heights. Wind direction correction was required for the Westermost Rough dataset due to known sensor alignment issues and changes in configuration during different measurement periods. Based on guidance from the data provider (Ørsted) and cross-validation with reference trends, consistent directional offsets of +4 degrees and -129 degrees were identified and applied accordingly. Anholt data remained consistent and did not

require correction. Since the raw data contains inconsistencies, preprocessing and filtering were necessary to ensure accurate input features for wind power prediction. The SCADA data has a temporal resolution of 10 minutes, while the LiDAR data is recorded at 1-minute intervals. To align both datasets, LiDAR measurements were aggregated to 10-minute intervals using statistical features (mean, min, max, and standard deviation). This ensured consistency between data sources and avoided time misalignment in the training process. For model development, the dataset was split chronologically into 80% training and 20% testing subsets. The split was done after preprocessing, with care taken to prevent temporal overlap that could lead to data leakage.”

3. Methods

To leverage the insights from these datasets, we employ advanced machine learning techniques, namely XGBoost and Bi-LSTM, to improve short-term wind power predictions. The forecast horizon used in this study is 1 hour ahead, consistent with short-term wind power prediction objectives. Although the raw LiDAR and SCADA data are sampled at high frequency, engineered features enable the models to learn temporal dependencies that are meaningful over a 30-minute to 6-hour horizon. For both models, we selected a relevant subset of SCADA and LiDAR parameters, focusing on wind speed (mean, min, max), wind direction, turbulence intensity, and temperature. These were further processed to generate lagged features and rolling window statistics (mean, standard deviation, sum) for 3, 6, and 12 time steps, as described in Appendix B. XGBoost then performed its internal feature selection. At the same time, Bi-LSTM used all engineered features directly. The implementations of XGBoost and Bi-LSTM models used in this study are based on established Python libraries. Specifically, the XGBoost model was developed using the `xgboost` [13] Python package, while the Bi-LSTM model was implemented using TensorFlow and Keras frameworks [14]. Although the architectures were tailored for this study’s forecasting objectives, the core components of each algorithm followed standard library implementations. Relevant references to these libraries have been added.

3.1. XGBoost

Extreme Gradient Boosting (XGBoost) is a widely used machine learning algorithm for classification and regression tasks, known for its efficiency and scalability. It is widely used for classification and regression tasks due to its efficiency, scalability, and ability to handle structured datasets effectively. XGBoost operates by constructing multiple decision trees sequentially, where each tree attempts to correct the errors made by the previous trees. The final prediction is obtained by aggregating the results from all trees using weighted averaging or majority voting.

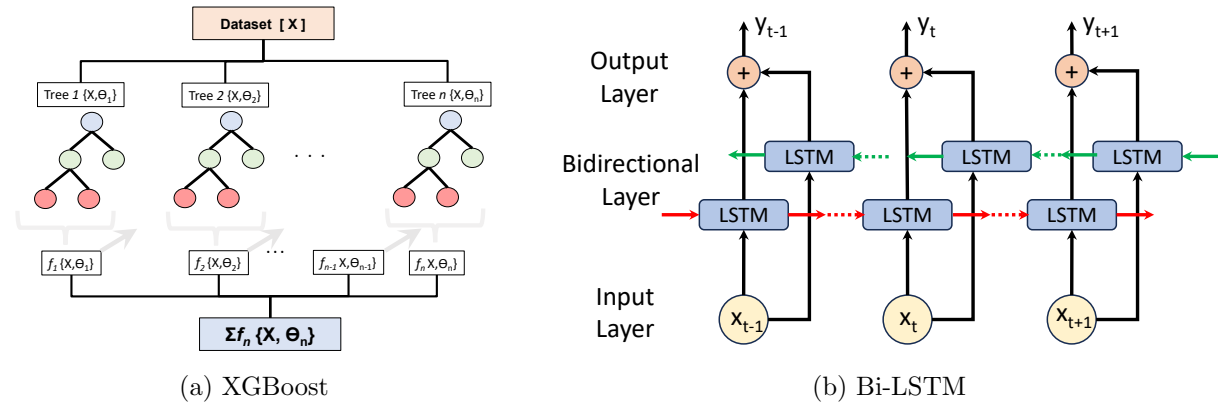


Figure 2: Schematic representation of the machine learning models. Depicting the iterative boosting process for XGBoost and illustrating the flow of information in both forward and backward directions for Bi-LSTM.

Figure 2(a) illustrates the fundamental concept behind XGBoost, where multiple weak learners (decision trees) are iteratively trained on the dataset. Each tree is built based on the residual errors of the previous trees, gradually improving model accuracy. The algorithm assigns different weights to individual trees, optimizing the overall performance using an objective function that minimizes error while incorporating regularization techniques to prevent overfitting. A key advantage of XGBoost is its ability to efficiently handle missing data, feature selection, and large-scale datasets. Additionally, it incorporates parallel processing, tree pruning, and L1/L2 regularization to enhance model generalization. In wind power prediction, XGBoost can learn from complex, high-dimensional data by capturing non-linear relationships between wind characteristics and power generation.

3.2. Bi-LSTM

A bidirectional long short-term memory (Bi-LSTM) network, shown in Figure 2(b), is an extension of the traditional LSTM architecture that improves the model's ability to learn from sequential data by processing information in both forward and backward directions. Unlike standard LSTM, which only retains past dependencies, Bi-LSTM consists of two LSTM layers: one that processes the input sequence in the forward direction and another in the reverse direction. This structure enables the network to capture both past and future context simultaneously, enhancing performance in time-series prediction and sequence modeling tasks. The forward layer receives the input sequence in its original order, while the backward layer processes the sequence in reverse. The outputs of both layers are then concatenated or combined before passing through the subsequent layers of the network. This bidirectional mechanism allows the network to extract richer contextual information, making it particularly useful in wind power prediction, speech recognition, and natural language processing applications. By leveraging past and future dependencies, Bi-LSTM enhances predictive accuracy, especially in cases where sequential dependencies are not strictly unidirectional. A summary of the model implementation is given in Appendix 'B'.

4. Results and discussion

4.1. Data filtering

The raw datasets from the Anholt and Westermøst Røgh wind farms contain missing values and outliers due to various external factors, including sensor malfunctions, adverse weather conditions, and transmission errors. A rigorous data filtering and preprocessing approach

was necessary to ensure the reliability of data for wind power prediction. The LiDAR and SCADA datasets include multiple measurement parameters recorded over extended periods, often containing gaps or erroneous readings. Interpolation techniques were applied to address missing values, including linear and nonlinear methods, ensuring continuity in the time-series data. Outlier detection methods were also employed to identify and remove anomalous values that could distort the predictive models. The data underwent normalization to scale input features to a uniform range, reducing bias in model training. Furthermore, wind direction measurements from Westermost Rough required corrections based on predefined offset values to ensure consistency across datasets. Through these preprocessing steps, we transformed raw wind farm data into a structured format suitable for machine learning models, improving prediction accuracy while preserving the integrity of wind characteristics essential for short-term prediction.

4.2. Wind power prediction

The power prediction for the Anholt wind farm (shown in Figure 3(a)) using XGBoost demonstrates a strong alignment between the predicted and actual values over the observed time series. The predicted power output follows the actual fluctuations, capturing the general trend with minimal deviation. However, some discrepancies can be observed during rapid transitions in power output, where the model slightly lags or overestimates specific peaks. This suggests that while XGBoost effectively learns the nonlinear relationships in wind power generation, further tuning or feature engineering may improve predictive accuracy in transient conditions. The power prediction for the Westermost Rough wind farm, shown in Figure 3(b), also closely matches predicted and true values, with the model successfully capturing overall variations in wind power output. Similar to Anholt, the model performs well in stable conditions but shows minor deviations during sharp power changes. The data was normalized according to the requirements of the data provider, Ørsted, to comply with privacy regulations and ensure the secure handling of sensitive information.

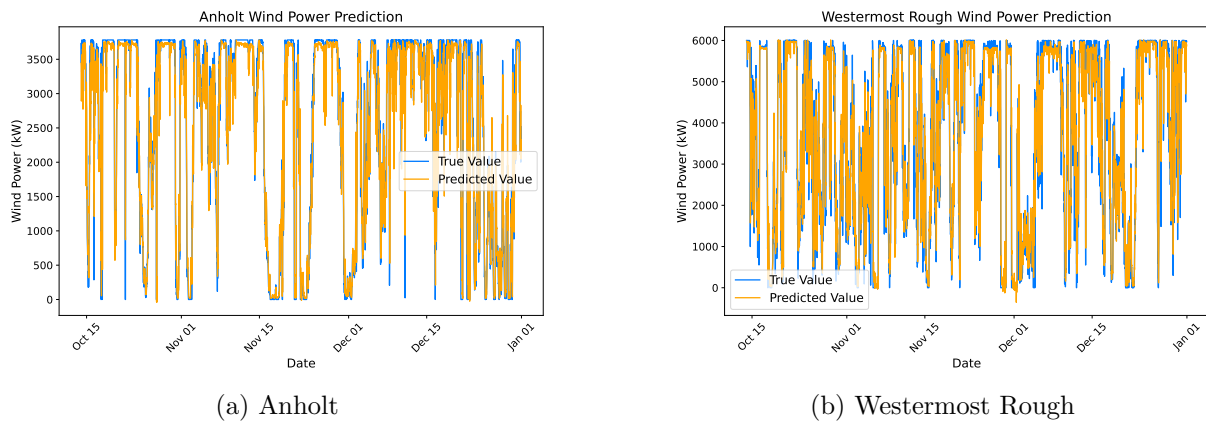


Figure 3: Power prediction using XGBoost model (a) Anholt wind farm (b) Westermost Rough wind farm. The predicted power output is compared against the true values recorded by the SCADA system.

The Bi-LSTM model for the Anholt wind farm (Figure 4(a)) shows strong predictive performance, closely tracking true wind power values and capturing general trends. However, minor deviations occur during rapid transitions due to its reliance on sequential dependencies. Similarly, for Westermost Rough (Figure 4(b)), Bi-LSTM effectively learns temporal patterns, though discrepancies arise during sharp power changes, suggesting a need for additional features or hybrid modeling. Additionally, a systematic underestimation is observed at high power

levels, particularly near peak output. This can be attributed to the model's training dynamics with MSE loss, which tends to prioritize accuracy across the full data range, often at the expense of extremes. Wang et al. [2] also observed that RNN-based models such as LSTM may systematically underestimate extreme output values due to the smoothing bias introduced by the MSE loss, which aligns with our observation. Compared to XGBoost, Bi-LSTM better captures long-term dependencies, while XGBoost excels in feature selection and fast computation. However, Bi-LSTM is more computationally intensive.

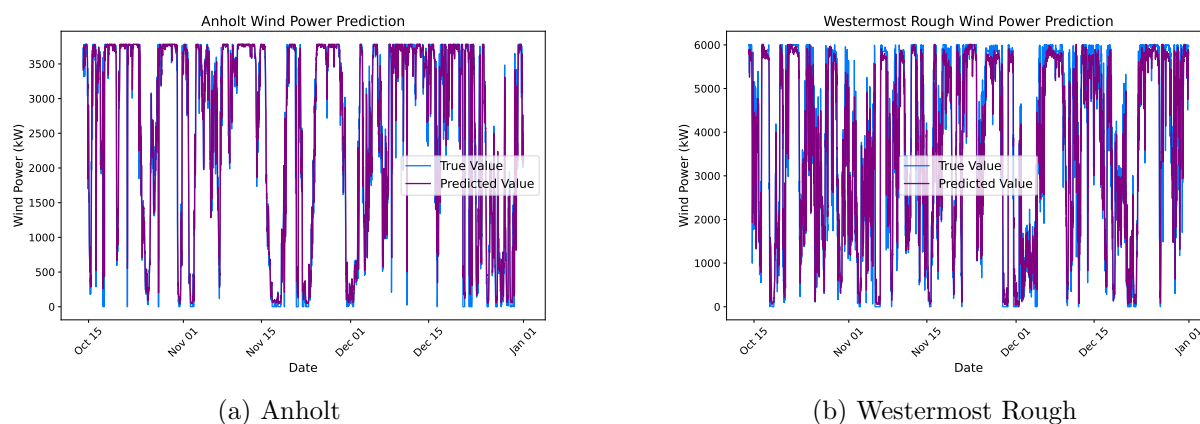


Figure 4: Power prediction using Bi-LSTM model (a) Anholt wind farm (b) Westernmost Rough wind farm

4.3. Model performance

Figure 5(a) compares the training time for Bi-LSTM and XGBoost models across both wind farms. Bi-LSTM takes significantly longer to train, with Anholt and Westernmost Rough models requiring 77.54 s and 45.33 s, respectively.

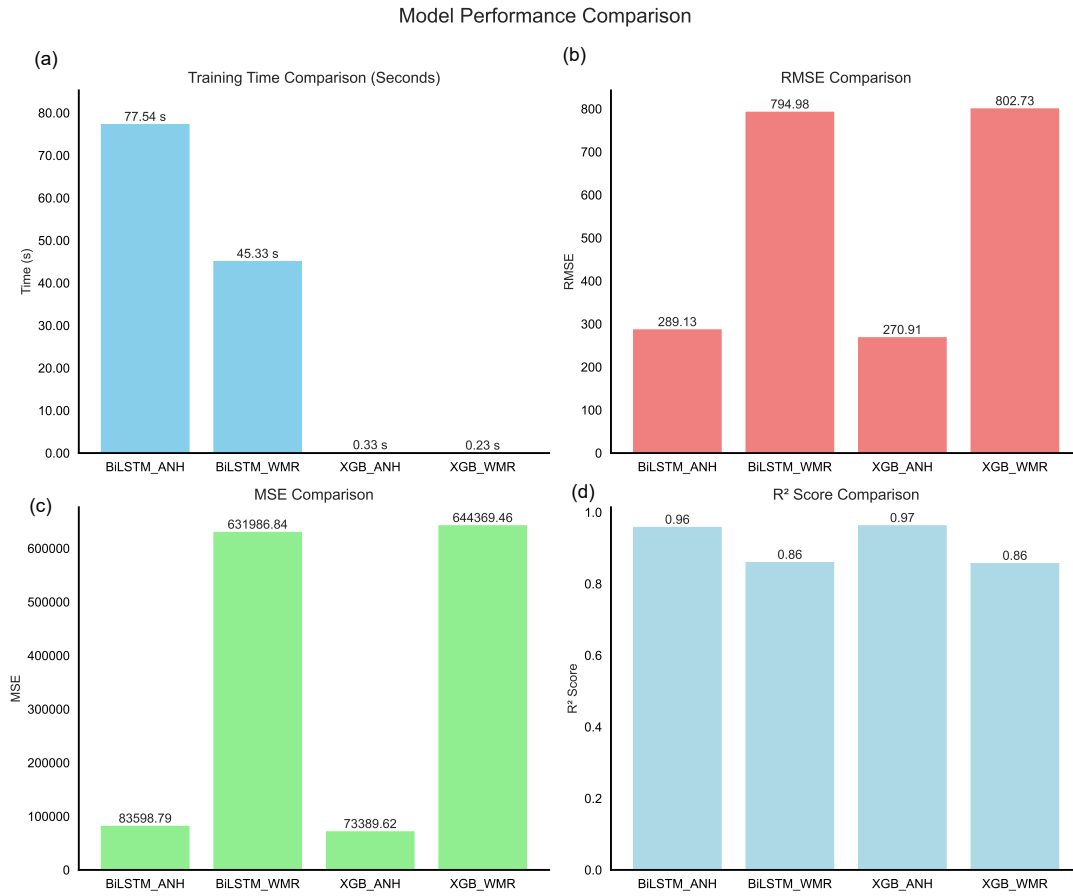


Figure 5: Model performance metrics across Bi-LSTM and XGBoost for Anholt and Westernmost Rough wind farms (a) Training time (b) RMSE comparison (c) MSE comparison (d) R^2 score comparison

In contrast, XGBoost is computationally efficient, with training times of only 0.33 s and 0.23 s. This highlights Bi-LSTM's complexity in handling sequential dependencies, whereas XGBoost provides faster model training, which is beneficial during model development and tuning. While real-time applications primarily rely on inference speed, short training times enable efficient experimentation, hyperparameter optimization, and potential periodic retraining in operational pipelines. Figure 5(b) presents the Root Mean Square Error (RMSE) values, where lower values indicate better predictive accuracy. For Anholt, Bi-LSTM achieves 289.13, slightly higher than XGBoost's 270.91. However, for Westernmost Rough, both models perform worse, with Bi-LSTM at 794.98 and XGBoost at 802.73. The performance degradation at Westernmost Rough suggests increased variability impacting prediction accuracy. Mean Square Error (MSE) results (shown in Figure 5(c)) are consistent with the RMSE trends. Bi-LSTM achieves an MSE of 83,598.79 at Anholt, slightly higher than XGBoost's 73,389.62. At Westernmost Rough, MSE is significantly larger for both models, with Bi-LSTM at 631,986.84 and XGBoost at 644,369.46, indicating higher prediction uncertainty at this site. Figure 5(d) illustrates the R^2 scores, where values closer to 1 indicate better model performance. For Anholt, Bi-LSTM achieves an R^2 of 0.96, while XGBoost slightly outperforms at 0.97. However, Westernmost Rough's models perform similarly, with an R^2 of 0.86, indicating moderate predictive capability.

Figure 6 presents the residual distributions for Bi-LSTM and XGBoost models at the Anholt and Westernmost Rough wind farms. The residuals represent the differences between

predicted and actual wind power values, with a more concentrated and symmetric distribution around zero, indicating better model performance. For Anholt, both Bi-LSTM (Figure 6(a)) and XGBoost (Figure 6(c)) show relatively symmetric residual distributions, suggesting good predictive performance with minimal systematic bias. However, Bi-LSTM appears to have slightly lower dispersion, indicating better error control. For Westernmost Rough, Bi-LSTM (Figure 6(b)) and XGBoost (Figure 6(d)) exhibit a wider spread of residuals, reflecting higher prediction uncertainties. The larger residual range for both models suggests greater challenges in predicting wind power at this site.

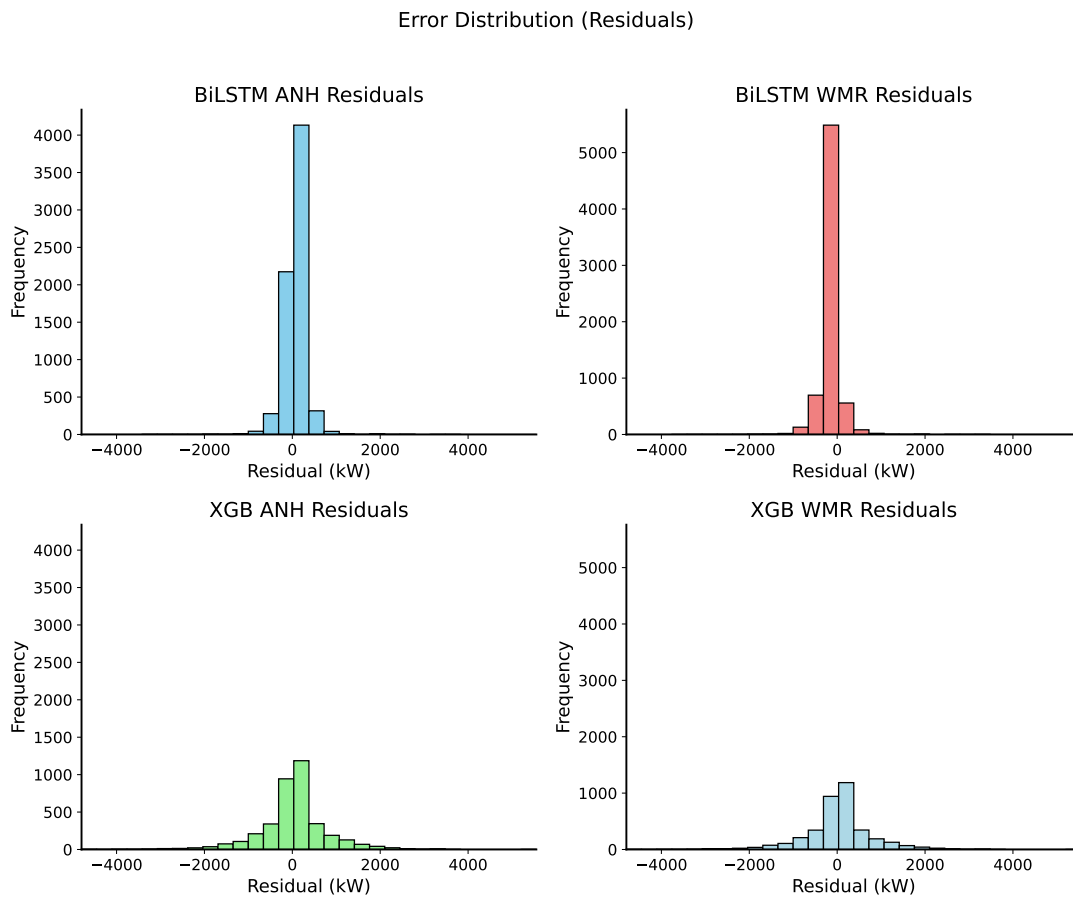


Figure 6: Residual distribution plots for Bi-LSTM and XGBoost models applied to the Anholt and Westernmost Rough wind farms. (a) Bi-LSTM at Anholt (b) Bi-LSTM at Westernmost Rough (c) XGBoost at Anholt (d) XGBoost at Westernmost Rough.

4.4. Wind speed loss

Accurate feature representation is critical for predictive modeling, particularly when integrating multi-source datasets such as LiDAR and SCADA. To assess data consistency and identify potential biases, we analyzed wind speed differences (SCADA @90 m – LiDAR @85.6 m LAT) across individual turbines in two selected rows at the Anholt wind farm. The results, shown in Figure 7, highlight the spatial variability of wind speed discrepancies within the farm.

In Row 1 (Figure 7 (a)), significant variation in mean wind speed differences is observed. Turbines such as A01, G01, and J01 exhibit the highest positive differences, reaching up to 0.86 m/s, suggesting that SCADA measurements reported consistently higher wind speeds than the LiDAR reference at these locations. Conversely, turbines like I01 and F01 show

negative differences, with I01 reaching -0.27 m/s, indicating local flow anomalies or potential underestimation by SCADA data. Overall, the spread ranges from approximately -0.27 m/s to 0.86 m/s, underscoring localized effects such as wake interactions, terrain-induced flow variability, or measurement misalignments.

In Row 2 (Figure 7 (b)), fewer turbines were analyzed, yet the trend of spatial variability persists. While turbine A17 shows a small positive bias (0.24 m/s), turbines C12, F11, and P13 display significant negative differences, with C12 experiencing the largest drop at -0.68 m/s. This consistent under-reporting by SCADA relative to LiDAR in this row could be indicative of complex flow structures or positioning effects relative to the LiDAR device. These observations highlight the importance of turbine-specific analysis when integrating LiDAR and SCADA datasets. The presence of both positive and negative differences across the farm emphasizes the influence of local turbulence, flow obstructions, and instrumentation sensitivity. Such insights are crucial for enhancing model input quality and, ultimately, the reliability of short-term power predictions.

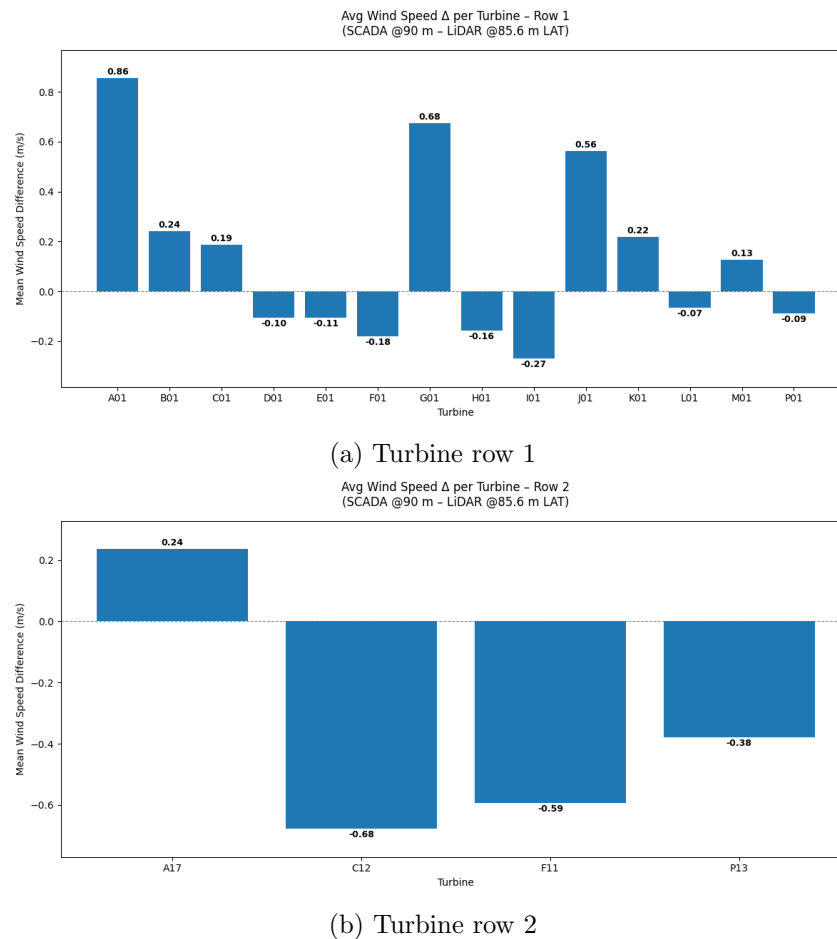


Figure 7: Wind speed loss analysis for two selected turbine rows in the Anholt Offshore Wind Farm, comparing LiDAR and SCADA measurements (a) Turbine row 1 (b) Turbine row 2

5. Conclusions

This study explored the application of data-driven models, specifically XGBoost and Bi-LSTM, for short-term wind power prediction using LiDAR and SCADA datasets from the Anholt and Westernmost Rough offshore wind farms. The results demonstrated that both models effectively

captured wind power variations, with XGBoost excelling in computational efficiency and Bi-LSTM better handling sequential dependencies. XGBoost achieved high predictive accuracy with lower training time, making it suitable for real-time applications. Bi-LSTM, leveraging its bidirectional architecture, provided improved performance in learning temporal dependencies but required significantly higher computational resources. These findings emphasize the variability in wind speed loss across different turbine locations, influenced by local turbulence and measurement uncertainties. The observed discrepancies between LiDAR and SCADA data can introduce noise and bias in model training, reducing prediction accuracy. By improving data consistency and optimizing model hyperparameters, the prediction model learns more reliable patterns. This leads to more accurate short-term forecasts, enabling better turbine scheduling, reduced energy loss, and improved overall wind farm efficiency. This ongoing research aims to integrate more variables and develop hybrid deep learning approaches to improve offshore wind power prediction accuracy and efficiency.

Appendix: A - Comparative table of key parameters for the Anholt and Westernmost Rough wind farms:

Parameter	Anholt	Westernmost Rough
Location	Kattegat, Denmark	North Sea, UK, 8 km off Yorkshire coast
Number of Turbines	111	35
Turbine Capacity (MW)	3.6	6
Total Capacity (MW)	400	210
Rotor Diameter (m)	120	154
Hub Height (m)	81.6	102
Wind turbine foundation	Monopile	Monopile
Water Depth (m)	15-19	16 – 20
Distance to Shore (km)	15	10
Commissioning	Summer 2013	Q1 2015

Appendix: B - Summary of Models:

XGBoost Model

- XGBoost regressor with:
 - 100 estimators.
 - Learning rate of 0.1.
 - Maximum depth of 5.
 - Random state set to 42.
 - Custom weighted MSE loss function.

Feature Engineering

- Lagged features up to 12 time steps for wind speed and power data.
- Rolling window statistics (mean, standard deviation, sum) for 3, 6, and 12 time-step windows.
- These engineered features allow the model to leverage historical trends and fluctuations over time windows that span 30 minutes to several hours, thus addressing the short-term prediction goal despite the high-frequency nature of the original input data.

Bi-LSTM Model

- A bidirectional LSTM model with three stacked LSTM layers.
- First layer: 128 units, ReLU activation, return sequences enabled.
- Second layer: 64 units, ReLU activation, return sequences enabled.
- Third layer: 32 units, ReLU activation.
- Two dropout layers (0.2) for regularization.
- Dense layer with 64 neurons and ReLU activation.
- Output layer with a single neuron using ReLU activation.
- Optimized with Adam and a custom weighted MSE loss function.

Acknowledgements

The authors would like to acknowledge Ørsted for providing the wind farm datasets.

References

- [1] Abdulgader Alsharif and Majdi Omar Ajaj. Harnessing offshore wind energy: Challenges and technological advancements. *The Open European Journal of Engineering and Scientific Research (OEJESR)*, pages 12–22, 2025.
- [2] Yun Wang, Runmin Zou, Fang Liu, Lingjun Zhang, and Qianyi Liu. A review of wind speed and wind power forecasting with deep neural networks. *Applied energy*, 304:117766, 2021.
- [3] Mike Optis and Jordan Perr-Sauer. The importance of atmospheric turbulence and stability in machine-learning models of wind farm power production. *Renewable and Sustainable Energy Reviews*, 112:27–41, 2019.
- [4] Dae-Young Kim, Yeon-Hee Kim, and Bum-Suk Kim. Changes in wind turbine power characteristics and annual energy production due to atmospheric stability, turbulence intensity, and wind shear. *Energy*, 214:119051, 2021.
- [5] Mathieu Pichault, Claire Vincent, Grant Skidmore, and Jason Monty. Short-term wind power forecasting at the wind farm scale using long-range doppler LIDAR. *Energies*, 14(9):2663, 2021.
- [6] Jannis Tautz-Weinert and Simon J Watson. Using SCADA data for wind turbine condition monitoring—a review. *IET Renewable Power Generation*, 11(4):382–394, 2017.
- [7] Mehdi Neshat, Meysam Majidi Nezhad, Ehsan Abbasnejad, Seyedali Mirjalili, Lina Bertling Tjernberg, Davide Astiaso Garcia, Bradley Alexander, and Markus Wagner. A deep learning-based evolutionary model for short-term wind speed forecasting: A case study of the lillgrund offshore wind farm. *Energy conversion and management*, 236:114002, 2021.
- [8] Coleman Moss, Romit Maulik, and Giacomo Valerio Iungo. Augmenting insights from wind turbine data through data-driven approaches. *Applied Energy*, 376:124116, 2024.
- [9] Jiaqing Yin, Jieyi Ding, Yang Yang, Jie Yu, Lu Ma, Wenhao Xie, Debang Nie, Musa Bashir, Qianni Liu, Chun Li, et al. Wave-induced motion prediction of a deepwater floating offshore wind turbine platform based on Bi-LSTM. *Ocean Engineering*, 315:119836, 2025.
- [10] Haris Hameed Mian, Muhammad Salman Siddiqui, Arvind Keprate, and Sathyajith Mathew. Predictive modeling of semi-submersible floater motion using Bi-LSTM model. In *Journal of Physics: Conference Series*, volume 2875, page 012029. IOP Publishing, 2024.
- [11] Anping Wan, Chenyu Du, Wenbin Gong, Chao Wei, Khalil AL-Bukhaiti, Yunsong Ji, Shidong Ma, Fareng Yao, and Lizheng Ao. Using transfer learning and XGBoost for early detection of fires in offshore wind turbine units. *Energies*, 17(10):2330, 2024.
- [12] Daeyoung Kim, Geonhwa Ryu, Chaejoo Moon, and Bumsuk Kim. Accuracy of a short-term wind power forecasting model based on deep learning using LiDAR-SCADA integration: A case study of the 400-mw anholt offshore wind farm. *Applied Energy*, 373:123882, 2024.
- [13] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.
- [14] Abadi Martín, Agarwal Ashish, Barham Paul, Brevdo Eugene, Chen Zhifeng, Citro Craig, S Corrado Greg, Davis Andy, Dean Jeffrey, Devin Matthieu, et al. Tensorflow: Large-scale machine learning on heterogeneous systems. *Software available from tensorflow.org*, 7, 2015.



DSPACE

<https://dspace.org/>

Extracting data-driven insights from LiDAR and SCADA data for improved analysis of offshore wind farms

Mian, Haris Hameed; Rognskog Nordbak, Bjorn Eirik; Skoglund, Eirik; Asim, Taimoor; Yang, Liang ...

2025-10-01

IOP Publishing

<https://doi.org/10.1088/1742-6596/3131/1/012031>

Mian HH, Rognskog Nordbak BE, Skoglund E, et al., (2025) Extracting data-driven insights from LiDAR and SCADA data for improved analysis of offshore wind farms. Journal of Physics: Conference Series, Volume 3131, October 2025, Article number 012031