E3S Web of Conferences **680**, 00116 (2025) https://doi.org/10.1051/e3sconf/202568000116
ICEGC'2025

Advancing Wake Loss Prediction in Wind Farms through Analytical Machine Learning Hybridization

Bouchra Talbi[1], *Bellat* Abdelouahad[2], *Khalifa* Mansouri[1]
[1]Laboratory of Modelling and simulation of intelligent industrial systems (M2S2I), ENSET Mohammedia, Hassan 2 university, Casablanca, Morocco
[2]PCMT Laboratory, ENSAM Rabat, Mohammed V University in Rabat, Rabat 10000, Morocco

Abstract. Wake models are still important for wind farm evaluation because they are computationally efficient when predicting turbine interactions and performance of the wind farm layout. Although the Jensen model is the simplest of the wake models and is often used, its assumption of linear wake expansion limits its accuracy when modelling wake interactions in complex terrain. To overcome the limitation of the Jensen model, this work introduces a hybrid framework that combines the analytical Jensen model with Extreme Gradient Boosting (XGBoost). The proposed framework utilizes wake predictions based on physics from the Jensen model combined with operational SCADA data, with XGBoost serving as a residual correction layer that learns the errors of the Jensen model and corrects them to improve downstream wind speed predictions. The results of this framework indicate considerable improvements from the Jensen model the R^2 increased from 0.8228 to 0.9237, RMSE decreased from 1.4011 m/s to 0.9201 m/s, MAE decreased from 0.9699 m/s to 0.6759 m/s. These results demonstrate the efficacy and novelty of hybridized physics machine learning approaches as a scalable, interpretable solution for understanding wake loss in wind energy applications.

Keywords: Wake models, Hybrid Wake Modeling, Wind Farm Optimization, Machine Learning.

1 Introduction

The rising pressure for clean and sustainable energy has placed renewable sources at the forefront of global energy policies[1]. As one of the fastest growing renewable energy technologies, wind energy has become a vital pillar of energy transition, and climate mitigation and energy policy at a global level[2]. Wind is plentiful and clean at the point of use, and growth in global installed capacity has been exponential year-on-year[3]. Wind turbines are a significant part of modern power systems, and are contributing to reduced greenhouse gas emissions and reducing reliance on fossil fuels[4]. However, despite advances in capacity, the effective use of wind resources is challenging. Wind energy is variable, and the interaction of turbines in groups or large farms creates wake effects[5]. Upstream turbines extract momentum from the flow, which means that downstream turbines experience a reduction in wind speed, an increase in turbulence, and the likelihood of greater fatigue loads[6]. The effects of wake-induced phenomena can lead to significant loss of power (often in the range of 10–20%) in large cumulative onshore wind farms, and compromises on the economic or operational performance of wind energy[7]. In order to remedy this situation, a number of modeling approaches have emerged. First is high-fidelity computational approaches like Computational Fluid Dynamics (CFD), Large Eddy Simulations (LES), and Reynolds-Averaged Navier–Stokes (RANS) models that can quantitatively describe wake dynamics and turbulence evolution[8]. These techniques serve the important purpose of accurately describing wake impacts but will be limited to very few cases showing detailed wake interactions between turbine components, prohibitive computational expense (large computer clusters), and inherent simplicity (you cannot turn growth dynamics into a function for wind farm optimization). Secondly, analytical wake models are relatively inexpensive to compute and are used extensively in all aspects of wind energy in industry[9]. Specifically, the Jensen model has been used for decades in planning wind energy farms and estimating energy yield because of its straightforward nature[10]. Despite its accessibility and established example use, the limitations associated with the Jensen model are large you cannot assume linearity in the expansion of a wake for the total wind speed, it does not physically explain how a turbine allows waked air to have a velocity deficiency, or the relationship between variation from both topography and atmospheric changes. A recent trend in the literature shows increasing interest in hybrid approaches that retain the interpretive nature of physics-based models while introducing flexibility and nonlinear methods from learning. The idea can be summarized as follows: analytical models capture the physics of wake dynamics while data driven algorithms can model non-linear interactions that simple analytical models do not touch. In the context of wind energy, machine learning has already been applied to a variety of topics like power forecasting, fault detection and wake prediction. However, the potential for using ensemble learning algorithms for wake correction is still relatively unexplored. In this work, we present a hybrid structured approach that brings together the Jensen wake model with Extreme Gradient Boosting (XGBoost) a tree-based ensemble learning method popular for its strength and efficiency, and is able to capture nonlinear interactions between features. In our structured approach, Jensen based estimates are combined with upstream flow variables, and the inter-turbine spacing, while XGBoost acts as an additive correction layer to update the wake estimates. This design preserves the mathematical theory underpinning the method, but allows for enhancements due to data-driven corrections. We validate our proposed approach using one year of SCADA data collected from a 21turbine onshore wind farm in Scotland characterized by complex hilly terrain (Hill of Towie site). Overall, by combining physics-based models with advanced machine learning, our Jensen+XGBoost framework represents an interpretable, computationally efficient, and scalable approach for improving wind farm performance evaluations. This study introduces a new hybrid concept that combines an analytical model with a machine learning approach, aiming to link physical interpretability with the flexibility of data-driven prediction. In contrast to conventional analytical models that rely on simplified assumptions or machine learning approaches that often act as black boxes, the Jensen+XGBoost model brings together the advantages of both analytical and machine learning models. The analytical Jensen component preserves the theoretical foundations and physical interpretability of wake behavior. The XGBoost component learns the residual errors and nonlinear interactions that the classical model cannot represent.

Moreover, unlike previous work that centered primarily on power forecasting or simplified cases, this study directly corrected downstream wind speed predictions, using one full year of real SCADA data from the same onshore site in Scotland, characterized by complex hilly terrain. This coupling enables more accurate and generalizable modeling of wake effects without compromising computational efficiency thereby offering an original and scalable contribution to advancing hybrid physics-informed machine learning approaches for wind farm performance assessment. The rest of the paper is structured as follows. In Section 2 the state of the art is provided, with an assessment of previous research on wake modelling and hybrid approaches that integrate analytical, machine learning and other methodologies. In Section 3, the Jensen–XGBoost hybrid method for wind speed correction and the experimental design is described. In Section 4 the results of the analysis comparing the original Jensen model and the Jensen–XGBoost enhanced methodology, are presented and discussed. Last, in Section 5 some final conclusions are summarized and anticipated future work is described including the combination of additional machine learning techniques and application to other real-world wind farm datasets.

2 Theoretical background

2.1 Analytical wake models

Analytical wake models have historically been the foundation of both design and operational studies of wind farms because they have good approachability and require minimal computational demands[11]. Among the various kinds of analytical models, the Jensen model introduced in 1983, is by far the most commonly used[12]. The model's development was based on a linear expansion of the wake and a tophat velocity deficit profile which allows the modeler to quickly estimate the flow downstream approach of the turbines in even the largest arrays. The general approachability of the model also lends itself to early assessments, as well as commercially based assessments. However, these assumptions will consistently lead to systematic biases, especially when using the models in complex terrain, with basic assumptions concerning heterogeneous atmospheric conditions, and recognizing that the behavior of wakes is usually nonlinear. Several notable additions developed to address the aforementioned simplifying assumptions have occurred in subsequent years. Gaussian wake models, for example, discard the uniform profile in favor of velocity deficit distribution that is more representative of reality, and so-called momentum conserving variants that try to take into account the additional physical constraints on the wakes[13]. Regardless of the improvements to these frameworks, analytical specifications will always be constrained by the structure of the simplifications that simple analytical models impose in comparison to the complexity of the real world they will always find it difficult to reproduce highly accurate realizations of the interaction of wakes.

2.2 High-fidelity simulations

Computational Fluid Dynamics (CFD) methods provide another distinct avenue of modeling the wake which allows for direct wake structure that outperforms the capabilities of analytical models. In large eddy simulations (LES) turbulence structures are resolved on a very fine scale and in numerical models based upon Reynolds-averaged Navier–Stokes (RANS) equations averaged fields are completely resolved which provide a computational advantage with the absence of turbulence structures. Both of these methods provide solutions that significantly capture the complex flow physics such as wake meandering and turbulence generation, along with incorporating atmospheric stability effects that simplified analytical formulations cannot account for[14]. However, the affordable accuracy of CFD comes with significant trade-offs namely computational expense, time integrated methods and the level of expertise required to use CFD models appropriately. These disadvantages leave model development impractical in systematic big picture wind farm optimization and real time use because of required time and person-power, relegating their use primarily to academic research and verification.

2.3 Machine learning in wind energy

The growing availability of SCADA data has driven utilizing machine learning (ML) methods in the wind energy domain for forecasting, fault detection, and wake modeling[15]. These methods can harness the ability to represent nonlinear relationships and complex interactions, which classical models of wake processes cannot. Artificial Neural Networks (ANNs) have been commonly applied to approximate turbine performance[16], while Convolutional Neural Networks (CNN) have been assessed concerning spatial flow features[17]. Recurrent Neural Networks (RNN) and their expansions in Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) apply to model sequential dependencies in time series data, which are appealing for dynamics of wake behavior[18]. Ensemble approaches, such as Random Forests (RF) and Support Vector Machines (SV), have also found their way into wind domain applications because of robustness with heterogenous or noisy data. Despite the further development of ML methods, many neural and kernel-based approaches are computationally expensive, sensitive to hyperparameter tuning, and deficient by the lack of transparency and interpretability of the output function. This has increased the interest in studying ensemble learning in particular tree-based methods that balance predictive power, computational speed, and transparency. Extreme Gradient Boosting (XGBoost) is a tree ensemble approach that acknowledges non-linear relationships and attempts to mitigate overfitting through the use of regularization[19]. It is fast to train, scale well to large SCADA datasets, and provides explainability, via Feature Importance, which allows users to understand which variables influence the outcome variable. Within this manuscript, XGBoost is used as a corrective layer on top of the Jensen model predictions by assimilating physics-based estimates of the wake with a data-driven correction to enhance the accuracy of the forecasting, while retaining the computational efficiency.

3 Methodology

3.1 Data preparation

This study used one year (2023) of SCADA measurements from the Hill of Towie wind farm in Scotland. The wind farm has 21 turbines, with a rotor diameter of 82 m and a hub height of 59 m. The data consists of wind speed, wind direction, yaw angle, active power, and coordinates of the turbines. The data was thoroughly cleaned prior to analysis. Variables were standardized, timestamps reformatted, and any missing, corrupted or physically inconsistent records were deleted. The coordinates were also processed so that distances and bearings could be calculated to create upstream downstream turbine relationships. The resulting dataset is fully cleaned and structured in terms of linking upstream conditions and spatial descriptors with Jensen estimates according to downstream responses in tabular format and so could be readily used to train ensemble models XGBoost.

3.2 Proposed Model

3.2.1 Jensen model

The Jensen model, first proposed in the early 1980s, is still one of the most popularly used analytical wake models despite being extremely simple to conceptualize and requiring relatively little computational effort. It is based on a momentum balance in the wake, with the assumptions of a linearly expanding wake cone, and a uniform velocity deficit profile. The conservation of momentum in the wake can be expressed as:

$$\pi r_0^2 u_r + \pi(r_w^2 - r_0^2)u_0 = \pi r_w^2 u_w \tag{1}$$

where u_0 is the free stream wind speed, u_r the velocity immediately behind the rotor, u_w the mean wind speed at a downstream distance x, R the rotor radius, and $r_w$the wake radius at that location.
According to actuator disk theory, the velocity just behind the rotor is defined as:

$$u_r = u_0(1 - 2a) \tag{2}$$

with a representing the axial induction factor. The wake radius is assumed to grow linearly with distance as:

$$r_w(x) = R + kx \tag{3}$$

where k is the wake decay constant, which depends on surface roughness and atmospheric conditions and is typically calibrated for the site under study.
By combining these relations, the normalized velocity deficit at a distance x can be written as:

$$\alpha = 0.5\, ln\left(\frac{Z}{Z_0}\right) \tag{4}$$

with z the hub height and $z_0$the surface roughness length.
By combining these relations, the normalized velocity deficit at a downstream distance x can be written as:

$$\frac{\Delta u}{u_0} = \frac{2a}{\left(1+\frac{kx}{R}\right)^2}, \tag{5}$$

and the downstream velocity is given by:

$$u_w(x) = u_0\left(1 - \frac{2a}{\left(1+\frac{kx}{R}\right)^2}\right) \tag{6}$$

This approach can provide a reasonable estimate of velocity losses induced by wake effects and related losses in power. The assumptions of linear wake expansion and area within a uniform velocity deficit restrict its predictive ability in complex landscapes such as the Hill of Towie site.

3.2.2 Hybrid jensen+XGBoost

The Jensen model has a serious limitation in that its simple treatment of wake expansion assumes a uniform top-hat velocity deficit and also assumes the wake radius grows linearly. While it is simple and convenient, this formulation does not adequately replicate nonlinear wake interactions and the site-specific characteristics found at actual wind farms. To address

these limitations, we will introduce a hybrid framework which uses Extreme Gradient Boosting XGBoost to adjust the velocity predictions from our analytical model.
In this setting, the downstream wind speed is modeled as:

$$\widehat{u_w} = \mathcal{F}\left(u_0, d, u_{w,\text{Jensen}}; \theta_{\text{XGB}}\right) \quad (7)$$

The enhancement in predictive accuracy of the hybrid Jensen+XGBoost model is attributed to its residual learning mechanism. The Jensen model yields a physically interpretable baseline estimate of the downstream velocity, but it fails to capture nonlinear wake interactions, terrain effects, and atmospheric variations. The XGBoost model is trained to learn these residuals that is, the systematic differences between the SCADA measured wind speeds and the predictions from the Jensen model. By iteratively correcting the residuals, the model learns the complex relationships between upstream velocity, inter-turbine spacing, and local flow conditions that are not represented in the analytical formulation. The correction process significantly reduces prediction errors and increases correlation with observed data, resulting in a clear enhancement of the overall predictive accuracy while maintaining the interpretability and computational efficiency of the Jensen model.
where u_0 is the upstream velocity, d is the distance between turbines, $u_{w,\text{Jensen}}$ is the baseline estimate provided by the Jensen formulation, and θ_{XGB} denotes the parameters optimized by XGBoost.
The proposed methodology focuses on the downstream velocity itself and uses observed SCADA data to correct it, rather than modifying the wake radius. The XGBoost methodology builds trees sequentially and each tree improves the regression on the remaining errors from the previous iteration. In effect, this residual learning enhances the predictive skill of the hybrid model while retaining the physical interpretability of the Jensen baseline.
The training objective is defined as the minimization of the mean squared error between predicted and measured velocities

$$\mathcal{L} = \sum_{i=1}^{n}(y_i - \widehat{y_i})^2 \quad (8)$$

with y_i the observed downstream wind speeds from SCADA and $\widehat{y_i}$ the predictions of the hybrid framework. Model performance is evaluated using statistical indicators such as the coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute error (MAE), allowing comparison between the traditional Jensen model and the hybrid Jensen+XGBoost.

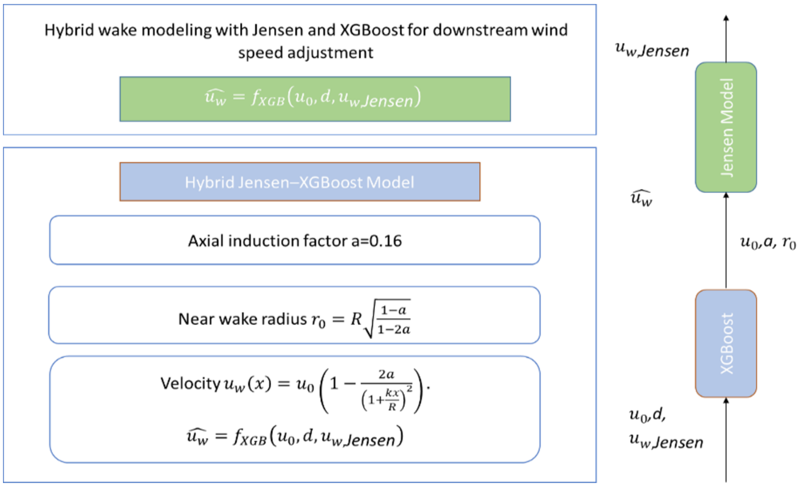

Fig. 1. Proposed Jensen–XGBoost hybrid methodology for wind speed adjustment.

4 Discussion

The comparison between the Jensen analytical framework and the Jensen + XGBoost hybrid representation provides very compelling evidence of the value of integrating machine learning into wake modeling. Using one full year of SCADA data from the Hill of Towie wind farm, the Jensen model baseline produced R^2=0.8228, RMSE = 1.4011 m/s, and MAE = 0.9699 m/these numbers highlight the well-known limitations of the model, the simplifying assumptions, notably uniform velocity deficits and linear geometric wakes limit the Jensen models utility in the context of real-world hilly terrain. As shown in Fig. 2, the hybrid method exhibited a clear ability to improve the downstream velocity estimate using a gradient boosting algorithm. The model achieved R^2 = 0.9237, RMSE = 0.9201 m/s, and MAE = 0.6759 m/s, ultimately capturing non-linear wake interactions more effectively and reducing prediction errors. Unlike sequential neural networks that tend to skew the wake geometry altogether, the XGBoost method was able to rely on tabular features, i.e., upstream velocity, inter-turbine distance, and Jensen estimates, to provide new estimates of downstream velocity that continued to minimize errors. The end resulting model is highly interpretable and practical. Fig. 3 additionally emphasizes this decrease in MAE and RMSE using the hybrid methodology instead of analytical methodology. As a whole, these results are not only indicative of the hybrid model's ability to interpolate the simplistic nature of a physics-based methodology with the flexibility of a data-driven methodology, but also encapsulates a significant real-world advantage to wind farm operators. Most importantly, being able to decrease our error metrics implies a more consistent depiction of wake losses and ultimately makes the model plausible to inform certain operational decisions and layout optimization tasks. It is essential to note there is still potential for error since we still rely on the accuracy of SCADA data and do not emphasize temporal dependencies; nevertheless, the hybrid framework exhibits also versatility and scalability potential, thus encouraging future exploration of the methodology for wind energy strategies in real-world applications.

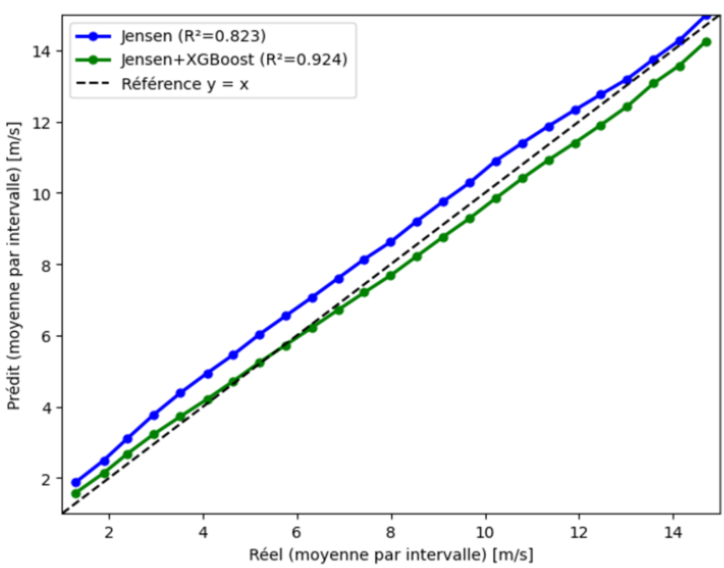

Fig. 2. Comparison between the Jensen Model and the Hybrid Jensen–XGBoost Model for Downstream Wind Speed Prediction.

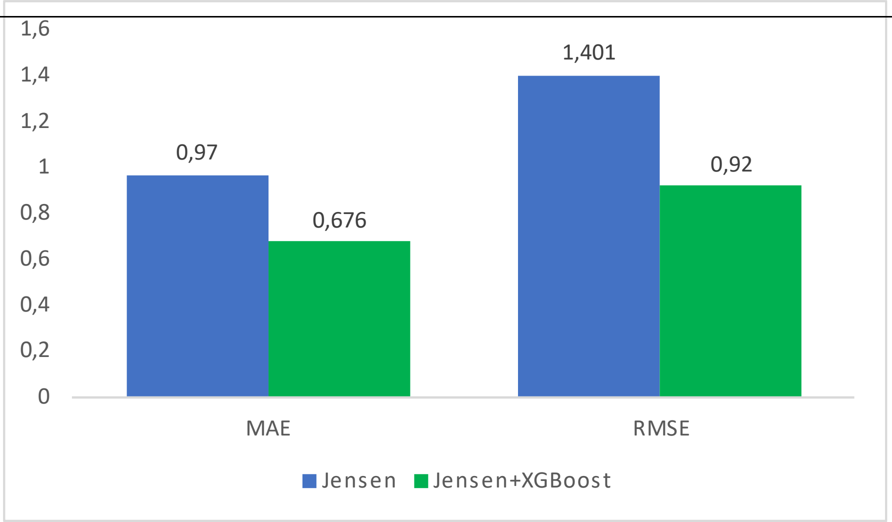

Fig. 3. Comparison of MAE and RMSE between the Jensen Model and the Hybrid Jensen–XGBoost Model.

5 Conclusions

Simulating wake interactions accurately is a main challenge of wind farm performance assessments, as downstream turbines are always at less wind speed and more turbulence because of turbines upstream. Analytical models, and especially the Jensen model, are still appealing to the industry because they are faster, but their simplified assumptions do not deliver measurable accuracy, particularly in heterogeneous terrains. Far better and more detailed, are CFD-based methods that simulate flows directly, but using CFD to analyze large scale or iterative simulation of flow is cost prohibitive. Thus, we are seeing hybrid methods that use physically based models and machine learning appear as a reasonable in between. In this paper, we suggested a Jensen+XGBoost hybrid model that was meant to simply correct the downstream wind speeds from Jensen. Our model used upstream wind speed, turbine spacing, and the identifiers that Jensen produces, with the XGBoost boosting algorithm modeling the relationships to correct predictions of wind speed using residual learning. We successfully applied the hybrid model to the Hill of Towie wind farm in Scotland, and the accuracy was significantly improved, where R^2=0.9237, RMSE = 0.9201 m/s, and MAE = 0.6759 m/s, compared to the baseline Jensen performance R^2=0.8228 , RMSE = 1.4011 m/s, MAE = 0.9699 m/s).The results validate that ensemble learning can usefully augment analytical wake models to make even more accurate predictions without sacrificing interpretability or adding computational expense. The hybrid framework therefore offers an exciting way to complement wind farm prediction tools in the future. Future works will apply the hybrid method to different terrain and turbine layout configurations, and explore hybridization through combinations of temporal architectures to better describe the dynamism of wind farm operations.

References

1. International Energy Agency, *World Energy Outlook 2023*. in World Energy Outlook. OECD, 2023. doi: 10.1787/827374a6-en.

2. P. Veers *et al.*, "Grand challenges in the science of wind energy," *Science*, vol. 366, no. 6464, p. eaau2027, Oct. 2019, doi: 10.1126/science.aau2027.
3. S. Tumse, M. Bilgili, A. Yildirim, and B. Sahin, "Comparative Analysis of Global Onshore and Offshore Wind Energy Characteristics and Potentials," *Sustainability*, vol. 16, no. 15, p. 6614, Aug. 2024, doi: 10.3390/su16156614.
4. Gondia Sokhna Seck, Bishal Parajuli, and X. G. Casals, "IRENA World energy transitions outlook 2023 volume 2," 2023, doi: 10.13140/RG.2.2.10445.65768.
5. L. J. Vermeer, J. N. Sørensen, and A. Crespo, "Wind turbine wake aerodynamics," *Progress in Aerospace Sciences*, vol. 39, no. 6–7, pp. 467–510, Aug. 2003, doi: 10.1016/S0376-0421(03)00078-2.
6. R. J. A. M. Stevens and C. Meneveau, "Flow Structure and Turbulence in Wind Farms," *Annu. Rev. Fluid Mech.*, vol. 49, no. 1, pp. 311–339, Jan. 2017, doi: 10.1146/annurev-fluid-010816-060206.
7. M. F. Howland, S. K. Lele, and J. O. Dabiri, "Wind farm power optimization through wake steering," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 116, no. 29, pp. 14495–14500, July 2019, doi: 10.1073/pnas.1903680116.
8. P. Veers *et al.*, "Grand Challenges: wind energy research needs for a global energy transition," *Wind Energ. Sci.*, vol. 7, no. 6, pp. 2491–2496, Dec. 2022, doi: 10.5194/wes-7-2491-2022.
9. M. Bastankhah and F. Porté-Agel, "A new analytical model for wind-turbine wakes," *Renewable Energy*, vol. 70, pp. 116–123, Oct. 2014, doi: 10.1016/j.renene.2014.01.002.
10. T. Göçmen, P. V. D. Laan, P.-E. Réthoré, A. P. Diaz, G. Chr. Larsen, and S. Ott, "Wind turbine wake models developed at the technical university of Denmark: A review," *Renewable and Sustainable Energy Reviews*, vol. 60, pp. 752–769, July 2016, doi: 10.1016/j.rser.2016.01.113.
11. L. Wang *et al.*, "Wind turbine wakes modeling and applications: Past, present, and future," *Ocean Engineering*, vol. 309, p. 118508, Oct. 2024, doi: 10.1016/j.oceaneng.2024.118508.
12. Niels Otto Jensen, "A note on wind generator interaction," Risø National Laboratory, Roskilde, Denmark, Technical Report Risø-M-2411, 1983. [Online]. Available: https://orbit.dtu.dk/en/publications/a-note-on-wind-generator-interaction
13. F. Blondel, "Brief communication: A momentum-conserving superposition method applied to the super-Gaussian wind turbine wake model," *Wind Energ. Sci.*, vol. 8, no. 2, pp. 141–147, Feb. 2023, doi: 10.5194/wes-8-141-2023.
14. M. Calaf, C. Meneveau, and J. Meyers, "Large eddy simulation study of fully developed wind-turbine array boundary layers," *Physics of Fluids*, vol. 22, no. 1, p. 015110, Jan. 2010, doi: 10.1063/1.3291077.
15. Z. Liu, H. Guo, Y. Zhang, and Z. Zuo, "A Comprehensive Review of Wind Power Prediction Based on Machine Learning: Models, Applications, and Challenges," *Energies*, vol. 18, no. 2, p. 350, Jan. 2025, doi: 10.3390/en18020350.
16. C. Yang, Z. Qian, Y. Pei, and L. Wei, "A Data-Driven Approach for Condition Monitoring of Wind Turbine Pitch Systems," *Energies*, vol. 11, no. 8, p. 2142, Aug. 2018, doi: 10.3390/en11082142.
17. J. A. Aird, E. W. Quon, R. J. Barthelmie, M. Debnath, P. Doubrawa, and S. C. Pryor, "Region-Based Convolutional Neural Network for Wind Turbine Wake Characterization in Complex Terrain," *Remote Sensing*, vol. 13, no. 21, p. 4438, Nov. 2021, doi: 10.3390/rs13214438.
18. Z.-H. Liu, C.-T. Wang, H.-L. Wei, B. Zeng, M. Li, and X.-P. Song, "A wavelet-LSTM model for short-term wind power forecasting using wind farm SCADA data," *Expert Systems with Applications*, vol. 247, p. 123237, Aug. 2024, doi: 10.1016/j.eswa.2024.123237.

https://doi.org/10.1051/e3sconf/202568000116

19. X. Xiong, X. Guo, P. Zeng, R. Zou, and X. Wang, “A Short-Term Wind Power Forecast Method via XGBoost Hyper-Parameters Optimization,” *Front. Energy Res.*, vol. 10, p. 905155, May 2022, doi: 10.3389/fenrg.2022.905155.