Environmental Data Science (0000), xx: 1–10
doi:10.1017/xxxx

CAMBRIDGE UNIVERSITY PRESS

APPLICATION PAPER

Estimating high-resolution profiles of wind speeds from a global reanalysis dataset using TabNet

Harish Baki[1]* and Sukanta Basu[2,3]

[1]Faculty of Civil Engineering and Geosciences, TU Delft, Delft, The Netherlands
[2]Atmospheric Sciences Research Center, University at Albany, Albany, New York, United States
[3]Department of Environmental and Sustainable Engineering, University at Albany, Albany, New York, United States
*Corresponding author. Email: h.baki@tudelft.nl

Received: ; **Revised:** ; **Accepted:**

Keywords: Chebyshev polynomials, Deep learning, Wind resource assessment

Abstract

The escalating demand for global wind power production, driven by the imperative need for sustainable energy sources, necessitates accurate estimation of vertical wind profiles for efficient wind turbine performance assessment. Traditional methods relying on empirical equations or similarity theory face limitations due to their applicability beyond the surface layer. Recent studies explore Machine Learning (ML) techniques to extrapolate wind speeds, but often focus on single levels, lacking a comprehensive approach to predict entire wind profiles. This study proposes a proof-of-concept in addressing the challenge, utilizing TabNet, an attention-based sequential deep learning model, to predict the entire wind profiles, provided by large-scale meteorological features from reanalysis. To make the methodology generic across datasets, the Chebyshev polynomials are used to approximate the wind profiles with Chebyshev coefficients. Trained on the meteorological features as inputs and the coefficients as targets, the TabNet better predicts unseen wind profiles for different wind conditions, such as high shear, low shear/well mixed, low level jet, and high wind, with a good accuracy. The methodology also addresses the correlation of wind profiles with associated atmospheric conditions by assessing the feature importance. The model demonstrates the feasibility of predicting wind profiles from large-scale meteorological variables, providing a valuable alternative to conventional methods.

Impact Statement

We applied deep learning in conjunction with Chebyshev polynomials to predict unseen wind profiles provided large-scale meteorological features, and understand the associated atmospheric conditions. The methodology can be extended to different locations and diverse wind profile datasets, and predicting wind profiles for several years in a round-robin manner.

1. Introduction

The demand for global wind power production has seen a substantial surge, driven by the growing recognition of renewable energy sources as a vital solution to combat climate change and the urgent need for a sustainable, low-carbon future (Nagababu et al., 2023). Although offshore wind power technology is still in its initial stages, it is predicted to grow rapidly, which is primarily attributed to offshore wind speeds being higher and more uniform as the distance from the coast increases (Guo et al., 2022).

Moreover, technological advancements have facilitated the deployment of the largest wind turbines to date, such as the MySE 16-260 with a 16 MW capacity (Atlas, 2023), having a rotor diameter of 260 m and a hub height of 152 m, making it the largest wind turbine to reach a towering height of 280 m.

For such a massive wind turbines, the traditional use of hub height wind speeds in estimating power output (IEC, 2005) is not suffice due to varying wind speeds across the rotor plane. Addressing this, the rotor equivalent wind speed approach considers a wind profile within the rotor swept area, enhancing reliability and accuracy in estimation of power output for such large wind turbines (Wagner et al., 2011; Van Sark et al., 2019). Furthermore, specific meteorological conditions lead to the formation of distinct wind profiles, such as low-level jets during strong stratification, well-mixed profiles during very unstable conditions, and Ekman profiles during neutral conditions (Durán et al., 2020). The atmospheric stability parameters like wind shear and turbulence intensity during these different wind profiles significantly influence power output (Elliott and Cadogan, 1990; Wharton and Lundquist, 2012). Beyond power output, wind profiles under different stability conditions, including shear, veer, and low-level jets, also contribute significantly to turbine loads (Dimitrov et al., 2015; Gutierrez et al., 2017; Park et al., 2015). These findings underscore the paramount importance of analyzing wind profiles in both wind resource assessment and turbine design analysis.

However, characterization of wind profiles across rotor-swept area has been hindered by the sparsity of observations at this height levels, since the deployment of wind measurements, such as wind masts and lidars are generally too expensive. There are many similarity theory-based and/or empirical equations exist, such as logarithmic law of the wall, Monin-Obukhov similarity theory, and Power law, to name a few, which have been used in wind resource assessment applications to extrapolate near-surface wind speeds from ground meteorological stations (Bañuelos-Ruedas et al., 2010) or satellites (Optis et al., 2021) to several vertical levels. However, these equations would often require additional information which are not usually measured by the ground stations, and would be only valid within the surface layer (Basu, 2023), while the contemporary turbines' swept-area lies far outside of this layer. At present, the only reliable approach for estimating wind profiles is to use mesoscale models. There have been several of such activities going on, such as, the Copernicus Regional Reanalysis for Europe (CERRA) (Schimanke et al., 2021), New European Wind Atlas (NEWA) (Hahmann et al., 2020; Dörenkämper et al., 2020), Dutch Offshore Wind Atlas (DOWA) (Wijnant et al., 2019), and Winds of the North Sea in 2050 (WINS50) (Dirksen et al., 2022), to name a few. However, measoscale model simulations require tremendous computational power.

In recent times, numerous machine learning (ML) studies have explored extrapolating near-surface wind speeds to rotor-swept heights. Mohandes and Rehman (2018) employed deep neural networks (DNN) to extrapolate wind speeds from lower lidar measurements to 120 m height, showing superior performance over the empirical local wind shear exponent method. Optis et al. (2021) investigated methods to extrapolate near-surface wind speeds from satellite-based wind atlases to hub heights, with ML models, particularly Random Forest (RF), outperforming traditional empirical methods. They highlighted that ML models trained on a limited number of lidars could accurately extrapolate winds at various surrounding locations. Building on this, Liu et al. (2023) used three ML (random forest) models to estimate wind speeds at 120 m, 160 m, and 200 m levels, incorporating large-scale weather features from ERA5 reanalysis and wind speed/direction from a remote sensing device. Including meteorological features significantly improved ML model accuracy compared to the empirical power law method. Yu and Vautard (2022) extended this approach, constructing RF and extreme gradient boosting (XGBoost) models to estimate a gridded dataset of 100 m wind speed using meteorological variables from the ERA5 reanalysis. However, these ML studies focused on specific extrapolation levels and lack generalization for entire vertical profiles.

2. Problem statement

In this paper, our objective is to explore a deep learning (DL) approach for wind profile estimation, leveraging large-scale features from existing reanalysis data. Traditionally, wind resource assessments involve collecting observations for one year using met-masts, sodars, and lidars and extrapolating winds for other years through the measure-correlate-predict (MCP) approach. To emulate this process, our focus is on training a DL model for one year and predicting for a different year. As a proof-of-concept, we utilized simulated high-resolution wind profiles from CERRA reanalysis and large-scale meteorological features from ERA5 (5th generation European Centre for Medium-Range Weather Forecasts (ECMWF)) (Hersbach et al., 2020) at a specific location. The challenge lies in generalizing the methodology for easy adoption with any datasets, could be observational data from various measurements.

To achieve this, we initially approximate the CERRA wind profile using Chebyshev polynomials, representing them with five coefficients. Using these coefficients as targets and ERA5 meteorological features as inputs, a DL model is trained. While the aforementioned machine learning models like Random Forest (RF) and XGBoost excel in regression problems, they are designed for predicting single targets. Given our objective of predicting all coefficients simultaneously to obtain collective wind profile information, we opt for state-of-the-art TabNet (Arik and Pfister, 2021), an attention-based sequential deep learning model. Additional details about the data used are provided in subsection 3.1, the Chebyshev coefficient estimation is explained in subsection 3.2, and the training procedure is outlined in subsection 3.3. The study's results are presented in section 4, with concluding remarks in section 5.

3. Data and Methodology

3.1. Data

3.1.1. Wind speed at different height levels

The CERRA is a state-of-the-art reanalysis developed by the collaborative efforts of the Swedish Meteorological and Hydrological Institute (SMHI), Norwegian Meteorological Institute (MET Norway), and Meteo-France. CERRA provides height level wind speed at 12 vertical levels, that are: 10, 15, 30, 50, 75, 100, 150, 200, 250, 300, 400, and 500 meters above sea level. The dataset exist as analysis at every 3rd hour and as forecast at lead hours of 1, 2, and 3. In the present study, we utilized the three-hourly analysis and the corresponding forecasts at lead hours of 1 and 2, collectively making an hourly dataset. We collected point-based time-series wind profiles from 0000UTC on 1st of Jan 2000 to 2300UTC on 31st of Dec 2001, at the FINO1 site (54.0143N, 6.58385E), where several wind meteorological studies have been conducted (Durán et al., 2020).

3.1.2. Meteorological variables

This study utilizes 34 large-scale meteorological variables from the publicly available and globally acclaimed ERA5 reanalysis data as drivers for wind profiles. Out of these, 25 variables are adopted based on the studies of Kartal et al. (2023), which are: $\mathbf{W}_{10}$, $\mathbf{W}_{100}$, α, $\mathbf{u}_*$, $\mathbf{W}^i_{p10}$, T_2, T_0, T_s, T_{d2}, P_0, $\mathbf{H}$, H_S, H_L, TCC, LCC, $CAPE$, CIN, $\bar{\epsilon}$, ΔT_1, ΔT_2, ΔT_3, $HRSin$, $HRCos$, $DYSin$, and $DYCos$. These variables have same naming convention and description as mentioned in Table 1 of Kartal et al. (2023). A comprehensive details of the remaining 9 variables are provided in Table 1.

Table 1. *Description of the large-scale meteorological variables adopted from the ERA5 reanalysis.*

Type	Variable	Equation	Description	Units
Derived	$\mathbf{W}_{10}$	$\sqrt{U_{10}^2 + V_{10}^2}$	Wind speed at 10 m a.g.l. computed from zonal and meridional components	m s^{-1}
Derived	$\mathbf{W}_{100}$	$\sqrt{U_{100}^2 + V_{100}^2}$	Wind speed at 100 m a.g.l. computed from zonal and meridional components	m s^{-1}
Derived	α	$\frac{\log \mathbf{W}_{100}/\mathbf{W}_{10}}{\log (100/10)}$	Power-law exponent of wind profile within 10–100 m a.g.l.	–
Derived	$\mathbf{W}_{975}$	$\sqrt{U_{975}^2 + V_{975}^2}$	Wind speed at 975hPa computed from zonal and meridional components	m s^{-1}
Derived	$\mathbf{W}_{950}$	$\sqrt{U_{950}^2 + V_{950}^2}$	Wind speed at 950hPa computed from zonal and meridional components	m s^{-1}
Derived	$\Delta\mathbf{W}_{975-100}$	$\mathbf{W}_{975} - \mathbf{W}_{100}$	Difference in wind speed between 975hPa level and 100 m level	m s^{-1}
Derived	$\Delta\mathbf{W}_{950-975}$	$\mathbf{W}_{950} - \mathbf{W}_{975}$	Difference in wind speed between 950hPa level and 957hPa level	m s^{-1}
Raw	$\mathbf{u}_*$		Friction velocity	m s^{-1}
Raw	$\mathbf{W}^i_{p10}$		Instantaneous wind gust at 10 m a.g.l.	m s^{-1}
Raw	T_2		Air temperature at 2 m a.g.l.	K
Raw	T_0		Skin temperature	K
Raw	T_s		Upper-level soil temperature	K
Raw	T_{d2}		Dew point temperature at 2 m a.g.l.	K
Raw	P_0		Mean sea level pressure	Pa
Raw	$\mathbf{H}$		Boundary layer height	m
Raw	h_{cb}		Cloud base height	m
Raw	H_S		Instantaneous surface sensible heat flux	W m^{-2}
Raw	H_L		Instantaneous moisture flux	Kg m^{-2} s^{-1}
Raw	TCC		Total cloud cover	–
Raw	LCC		Low-level cloud cover	–
Raw	CAPE		Convective available potential energy	J kg^{-1}
Raw	CIN		Convective inhibition	J kg^{-1}
Raw	$\bar{\epsilon}$		Energy dissipation rate in boundary layer	J m^{-2}
Raw	T_{975}		Air temperature at 975hPa	K
Raw	T_{950}		Air temperature at 950hPa	K
Derived	ΔT_1	$T_2 - T_0$	Difference in air and skin temperatures	K
Derived	ΔT_2	$T_0 - T_s$	Difference in skin and soil temperatures	K
Derived	ΔT_3	$T_2 - T_{d2}$	Temperature dew point spread	K
Derived	ΔT_4	$T_{975} - T_2$	Difference in temperatures between 975hPa and 2 m a.g.l.	K
Derived	ΔT_5	$T_{950} - T_{975}$	Difference in temperatures between 950hPa and 975hPa	K
Derived	HRSin	$\sin\left(\frac{2\pi \text{Hour}}{24}\right)$	Sine encoding of hours	–
Derived	HRCos	$\cos\left(\frac{2\pi \text{Hour}}{24}\right)$	Cosine encoding of hours	–
Derived	DYSin	$\sin\left(\frac{2\pi \text{Day}}{365}\right)$	Sine encoding of Julian days	–
Derived	DYCos	$\cos\left(\frac{2\pi \text{Day}}{365}\right)$	Cosine encoding of Julian days	–

3.2. Estimating Chebyshev coefficients

Chebyshev polynomials allow one to approximate a function with smallest error as follows (Mason and Handscomb, 2002):

$$U(z) = \sum_{n=0}^{\infty} C_n T_n(z) \tag{1}$$

Here, we want to approximate the wind speed U as a function of height z, with the combination of Chebyshev polynomials T_n multiplied by the corresponding coefficients C_n. The polynomials of the first kind can be estimated through recurrence relation as follows:

$$T_0(z) = 1 \tag{2}$$

$$T_1(z) = z \tag{3}$$

$$T_{n+1}(z) = 2zT_n(z) - T_{n-1}(z) \tag{4}$$

In this study, we employed fourth-order Chebyshev polynomials (Figure 1, 1st column). Once computed, these polynomials simplify the problem into a system of linear equations, allowing for coefficient estimation through equation solving or inverse matrix multiplication. The variable z is normalized between -1 and 1 in real data before estimating coefficients. For a wind profile with 12 vertical levels and Chebyshev polynomials of order 4 (T_0 to T_4), five Chebyshev coefficients (C_0 to C_4) are estimated, reducing the wind profile's complexity to five coefficients. If at all there exist different wind profile data (could be observations) up to 500 m level, not necessarily at the same vertical levels as stated here, still the coefficient estimation works well, making this methodology generic across different datasets.

T_0 represents a constant line, with C_0 approximating mean wind speed. Similarly, T_1 corresponds to a diagonal line, and C_1 approximates wind shear. The parabolic profile of T_2 is captured by C_2, representing curvature in the wind profile. These three coefficients are expected to capture a significant portion of the wind profile, while higher-order coefficients account for small-scale variations.

To illustrate the capability of Chebyshev coefficients, we compared four wind profiles from CERRA with their Chebyshev approximations (Figure 1). These profiles, namely high shear, low shear/well-mixed, low-level jet (LLJ), and high wind, selected from (Durán et al., 2020), are crucial for wind energy applications. The figures demonstrate the effective approximation of CERRA wind profiles by Chebyshev coefficients.

3.3. Experimental setup

Figure 2(a) illustrates a complete flowchart of the training procedure adopted in this study. First, the ERA5 predictor meteorological variables as inputs and the estimated Chebyshev coefficients as targets are stacked side-by-side as a tabular data. Next, the entire data of year 2001 is kept aside for testing purpose. Now, among the data of year 2000, randomly selected six consecutive days of each month are used for validation purpose. The remaining data is adopted for training the models. This ensures that the training and validation data covers seasonality given one year sample size. The data splitting strategy is illustrated in Figure 2(b). After splitting, there are 7056 samples in training, 1728 samples in validation, and 8760 samples in testing. After this, a min-max normalization function is constructed on the targets of training data, using which the targets of training and validation data are normalized.

The TabNet consists of several hyperparameters, in which we chose to tune n_d (width of decision prediction layer), n_{steps} (number of steps in the architecture), $N_{independent}$ (number of independent Gated Linear Units), n_{shared} (number of shared Gated Linear Units), and $gamma$ (coefficient for feature reusage). The readers are encouraged to go through the study of Arik and Pfister (2021) for more information about the architecture and the hyperparameters. Next, a random search is employed for tuning the model hyperparameters from the parameter spaces of n_d:[4,8,16], n_{steps}:[3,4,5], $n_{independent}$:[1,2,3,4,5], n_{shared}:[1,2,3,4,5], and $gamma$:[1.1,1.2,1.3,1.4]. With these parameters, the TabNet model is trained on the training data using the mean squared error (MSE) as loss function, and is evaluated on the validation data. Once the training is completed, the validation loss is estimated and the model is saved as an external file. After the hyperparameter tuning, the entire training procedure is looped for 10 ensembles beginning from the random train-validaiton splitting. For each ensemble, there will be one best model saved. The inner loop of hyperparameter tuning makes the model robust

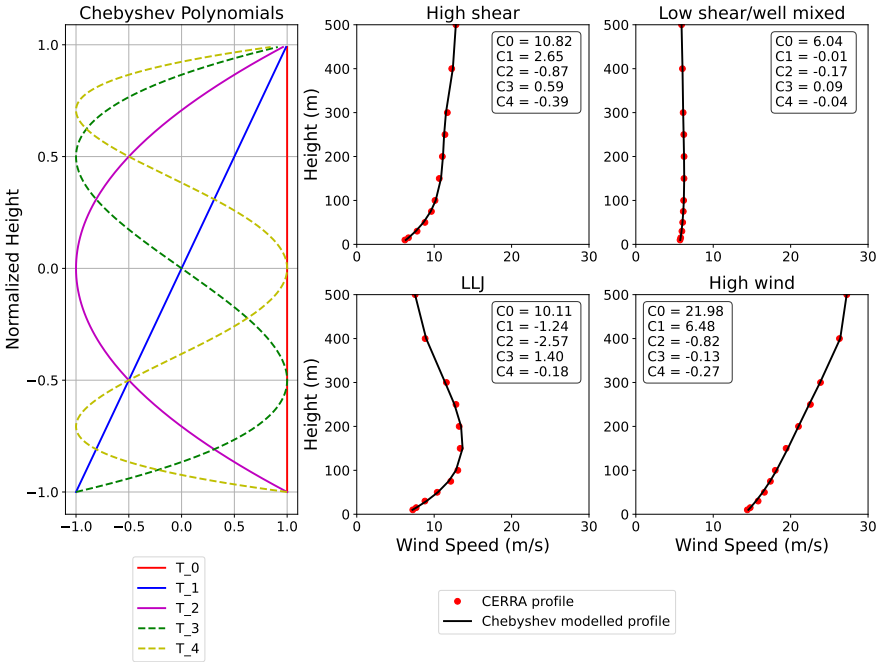

Figure 1. *Column 1: An illustration of Chebyshev polynomials of order $n = 4$ (T_0, T_1, T_2, T_3, and T_4), plotted against the normalized height $z = [-1, 1]$. Rest of the figures illustrates the vertical profiles of wind speed from CERRA and the Chebyshev approximated, for four well-known categories of wind regimes: high shear, low shear or well mixed, LLJ, and high wind..*

across the hyperparameters on the same data, while the outer loop is essential for ensemble model predictions. A sample learning curves obtained from one of the saved models is illustrated in Figure 2(c) which gives confidence that the models are training well.

4. Results

The model predictions are obtained for the test data, upon which the performance evaluation is conducted using metrics, namely mean absolute error (MAE), coefficient of determination (R^2), root mean square error (RMSE), and the mean absolute percentage error (MAPE). Figure 3(first column) illustrates a comparison of predicted coefficients from one of the ensemble models with respect to the test data using bivariate histograms. Among the coefficients, C_0 and C_1 are seen to have a large portion of occurrences falling on the diagonal line with a narrower spread. Notably, the model is seen to have highest predictability for C_0 with R^2 of 0.93 and moderate for C_1 with R^2 of 0.65. On the other hand, the remaining coefficients are seen to have values with high frequency of occurrence close to zero, while the values with low probability are showing a wider spread. For these coefficients, the predictability is not so good.

In addition, the influence of input meteorological variables on the predicted coefficients is also estimated based on the idea that the feature importance can be measured by looking at how much the score decreases when a feature is not available. From the feature importance, as shown in 3(second column), it is evident that the meteorological variables directly related to wind speed ($\mathbf{W}_{10}$ to $\mathbf{W}^i_{p10}$) are showing significant influence on the coefficients, which is expected. A major finding from the feature importance is that identification of atmospheric stability related variables, such as the boundary layer height ($\mathbf{H}$), instantaneous surface sensible heat flux (H_S), difference in air and skin temperatures (ΔT_1),

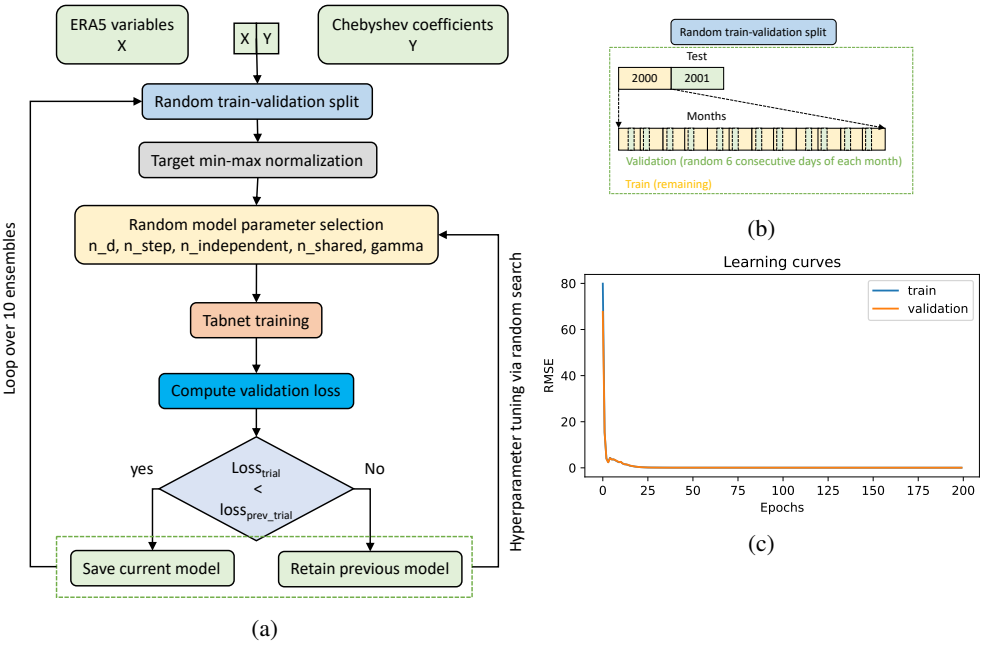

Figure 2. *a) Flowchart of the experimental setup used in this study to train the TabNet. b) Our strategy of splitting the entire dataset into train, validation, and test. c) Loss curves of one of the trained model, in which the train and validation RMSE values are plotted agains the training epochs..*

and the difference in 975hPa and 2 m air temperatures (ΔT_4), are also exerting significant influence on wind profiles. With these identified influential variables, one can accurately characterize different wind profiles, which would help in better estimation of wind power and their impact on wind turbines.

Since the main objective of this study is to predict the wind profiles given large-scale meteorological features from the ERA5, we wanted to see how well the predictions are coming from the trained models on test data. For that, the complete wind profiles are estimated using the predicted Chebyshev coefficients, and a sample of four profiles from the ten ensemble model predictions are presented in Figure 4. From the figure, it is evident that the predicted profiles are well matching with the CERRA test samples, even the dispersion between 10% to 90% and 25% to 75% are also very close to the median. However, not every prediction is turn out to be accurate, as shown in Figure 5, where wind profiles are illustrated during four different set of time instances. A keen observations of Figure 5 indicates that there exist a significant magnitude difference between the ERA5 and CERRA wind speeds, which could be the source of poor predictions at these instances. Nonetheless, the TabNet is well capable of capturing the wind profiles, such that the predicted profiles lie between the ERA5 and CERRA wind speeds. This uncertainty could be further reduced with training on large input samples and diversified locations. A possible future work is inevitable in this regard.

5. Conclusion

In this work, we demonstrated a proof-of-concept for estimating wind profiles using large-scale meteorological features from ERA5 reanalysis with the TabNet, an attention-based deep learning model. Chebyshev polynomials were employed to approximate wind profiles with 5 coefficients, ensuring the methodology's applicability across various wind profile datasets. Results indicated that TabNet effectively captured nonlinear dependencies between meteorological features and wind profiles across

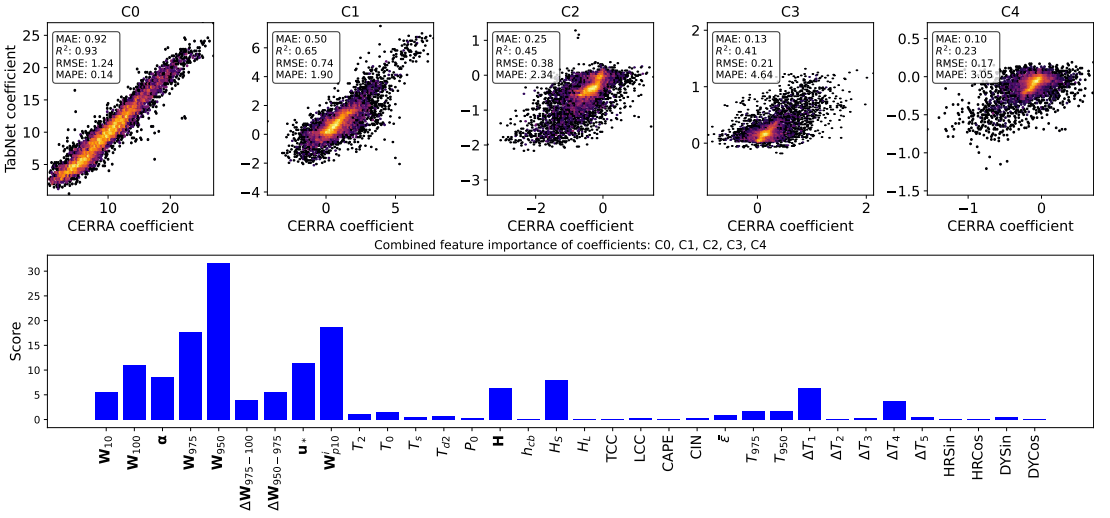

Figure 3. *First row: A comparison of Chebyshev coefficients (C_0, C_1, C_2, C_3 and C_4) between the test data and the model predictions using the Bivariate histograms. The probability of occurance is represented on a log scale with the color increasing from dark (low probability) to light (high probability). The evaluation scores, namely MAE, R^2, RMSE, and MAPE for each coefficient are provided in the text boxes. Second row: the combined feature importance of input variables on the target coefficients, on the test data..*

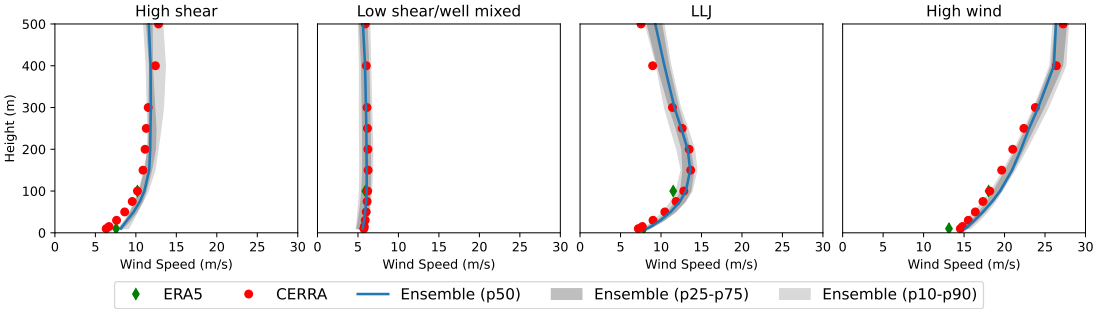

Figure 4. *A comparison of vertical profiles of wind speed from CERRA and the 10 ensemble model predictions, on four instances of test data for the selected wind regimes. Blue line represents the 50th percentile of ensembles, darker shade represents the ensemble between 25th and 75th percentiles, and the ligher shade represents the ensemble between 10th and 90th percentile. The wind speed from ERA5 at 10 m ($\mathbf{W}_{10}$) and 100 m ($\mathbf{W}_{100}$) are illustrated using green diamonds..*

different wind regimes. Feature importance analysis highlighted the significant influence of wind speed and atmospheric stability variables on coefficients.

However, there is room for improving model accuracy, as evidenced by the modest correlation in higher-order coefficients. Future work involves thorough parameter optimization, incorporating new meteorological variables, and exploring alternative deep learning models. Additionally, we plan to extend this methodology to different locations and diverse wind profile datasets, predicting for several years in a round-robin manner.

Funding Statement. This study was partially supported by the EU-SCORES and the Winds of the North Sea in 2050 (WINS50) projects. The EU-SCORES project is a part of the European Union's Horizon 2020 research and innovation programme under

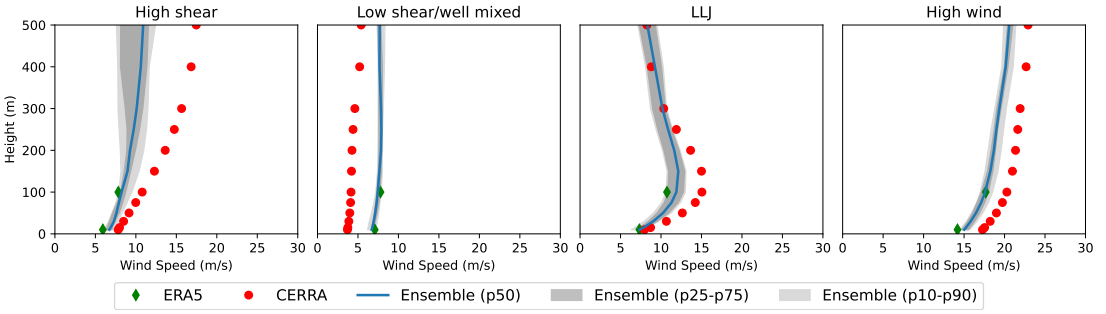

Figure 5. *Same as Figure 4, but a different set of time instances of the test data for the selected wind regimes..*

grant agreement No 101036457. The WINS50 project was sponsored by the Rijksdienst voor Ondernemend Nederland via the Top Sector Energy programme.

Competing Interests. None

Data Availability Statement. The ERA5 and CERRA reanalysis are downloaded from ECMWF CDO, available at https://cds.climate.copernicus.eu/cdsapp#!/search?type=dataset.

Ethical Standards. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

Author Contributions. Conceptualization: H.B; S.B. Methodology: H.B; S.B. Data curation: H.B. Data visualisation: H.B. Writing original draft: H.B; S.B. All authors approved the final submitted draft.

References

G. Nagababu, B. A. Srinivas, S. S. Kachhwaha, H. Puppala, S. V. A. Kumar, Can offshore wind energy help to attain carbon neutrality amid climate change? a gis-mcdm based analysis to unravel the facts using cordex-sa, Renewable Energy 219 (2023) 119400.

Y. Guo, H. Wang, J. Lian, Review of integrated installation technologies for offshore wind turbines: Current progress and future development trends, Energy Conversion and Management 255 (2022) 115319.

N. Atlas, World's largest wind turbine to feature 16 mw capacity, 2023. URL: https://newatlas.com/energy/worlds-largest-wind-turbine-myse-16-260/.

IEC, 12-1: Power performance measurements of electricity producing wind turbines, British Standard, IEC (2005) 61400–12.

R. Wagner, M. Courtney, J. Gottschall, P. Lindelöw-Marsden, Accounting for the speed shear in wind turbine power performance measurement, Wind Energy 14 (2011) 993–1004.

W. G. Van Sark, H. C. Van der Velde, J. P. Coelingh, W. A. Bierbooms, Do we really need rotor equivalent wind speed?, Wind Energy 22 (2019) 745–763.

P. Durán, S. Basu, C. Meißner, M. S. Adaramola, Automated classification of simulated wind field patterns from multiphysics ensemble forecasts, Wind Energy 23 (2020) 898–914.

D. L. Elliott, J. B. Cadogan, Effects of wind shear and turbulence on wind turbine power curves, Technical Report, Pacific Northwest Lab., Richland, WA (USA), 1990.

S. Wharton, J. K. Lundquist, Assessing atmospheric stability and its impacts on rotor-disk wind characteristics at an onshore wind farm, Wind Energy 15 (2012) 525–546.

N. Dimitrov, A. Natarajan, M. Kelly, Model of wind shear conditional on turbulence and its impact on wind turbine loads, Wind Energy 18 (2015) 1917–1931.

W. Gutierrez, A. Ruiz-Columbie, M. Tutkun, L. Castillo, Impacts of the low-level jet's negative wind shear on the wind turbine, Wind energy science 2 (2017) 533–545.

J. Park, L. Manuel, S. Basu, Toward isolation of salient features in stable boundary layer wind fields that influence loads on wind turbines, Energies 8 (2015) 2977–3012.

F. Bañuelos-Ruedas, C. Angeles-Camacho, S. Rios-Marcuello, Analysis and validation of the methodology used in the extrapolation of wind speed data at different heights, Renewable and Sustainable Energy Reviews 14 (2010) 2383–2391.

M. Optis, N. Bodini, M. Debnath, P. Doubrawa, New methods to improve the vertical extrapolation of near-surface offshore wind speeds, Wind Energy Sci. 6 (2021) 935–948.

S. Basu, Chapter 7 - vertical wind speed profiles in atmospheric boundary layer flows, in: T. M. Letcher (Ed.), Wind Energy Engineering (Second Edition), Academic Press, 2023, pp. 75–85.

S. Schimanke, M. Ridal, P. Le Moigne, L. Berggren, P. Undén, R. Randriamampianina, U. Andrea, E. Bazile, T. Bertelsen, P. Brousseau, P. Dahlgren, L. Edvinsson, A. El Said, M. Glinton, S. Hopsch, L. Isaksson, R. Mladek, E. Olsson, A. Verrelle, Z. Wang, CERRA sub-daily regional reanalysis data for Europe on pressure levels from 1984 to present, Technical Report, Copernicus Climate Change Service (C3S) Climate Data Store (CDS), 2021. doi:10.24381/cds.a39ff99f, accessed on 10-10-2023.

A. N. Hahmann, T. Sīle, B. Witha, N. N. Davis, M. Dörenkämper, Y. Ezber, E. García-Bustamante, J. F. González-Rouco, J. Navarro, B. T. Olsen, S. Söderberg, The making of the new european wind atlas – part 1: Model sensitivity, Geosci. Model Dev. 13 (2020) 5053–5078.

M. Dörenkämper, B. T. Olsen, B. Witha, A. N. Hahmann, N. N. Davis, J. Barcons, Y. Ezber, E. García-Bustamante, J. F. González-Rouco, J. Navarro, M. Sastre-Marugán, T. Sīle, W. Trei, M. Žagar, J. Badger, J. Gottschall, J. Sanz Rodrigo, J. Mann, The making of the new european wind atlas – part 2: Production and evaluation, Geosci. Model Dev. 13 (2020) 5079–5102.

I. Wijnant, B. van Ulft, B. van Stratum, J. Barkmeijer, J. Onvlee, C. de Valk, S. Knoop, S. Kok, G. Marseille, H. K. Baltink, et al., The dutch offshore wind atlas (dowa): Description of the dataset, Royal Netherlands Meteorological Institute, Ministry of Infrastructure and Water Management, De Bilt (2019).

M. Dirksen, I. Wijnant, A. Siebesma, P. Baas, N. Theeuwes, Validation of wind farm parameterisation in Weather Forecast Model HARMONIE-AROME: Analysis of 2019, Delft University of Technology, Netherlands, 2022.

M. A. Mohandes, S. Rehman, Wind speed extrapolation using machine learning methods and LiDAR measurements, IEEE Access 6 (2018) 77634–77642.

B. Liu, X. Ma, J. Guo, H. Li, S. Jin, Y. Ma, W. Gong, Estimating hub-height wind speed based on a machine learning algorithm: Implications for wind energy assessment, Atmospheric Chemistry and Physics 23 (2023) 3181–3193.

S. Yu, R. Vautard, A transfer method to estimate hub-height wind speed from 10 meters wind speed based on machine learning, Renewable and Sustainable Energy Reviews 169 (2022) 112897.

H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, et al., The era5 global reanalysis, Quarterly Journal of the Royal Meteorological Society 146 (2020) 1999–2049.

S. Ö. Arik, T. Pfister, Tabnet: Attentive interpretable tabular learning, in: Proceedings of the AAAI conference on artificial intelligence, volume 35, 2021, pp. 6679–6687.

S. Kartal, S. Basu, S. J. Watson, A decision tree-based measure-correlate-predict approach for peak wind gust estimation from a global reanalysis dataset, Wind Energy Science Discussions 2023 (2023) 1–25.

J. C. Mason, D. C. Handscomb, Chebyshev polynomials, CRC press, 2002.