



A Two-Step Spatio-Temporal Framework for Turbine-Height Wind Estimation at Unmonitored Sites from Sparse Meteorological Data

Eamonn Organ¹, Maeve Upton¹, Denis Allard², Lionel Benoit², and James Sweeney¹

¹Department of Mathematics and Statistics, University of Limerick, Limerick, Ireland

²Biostatistics and Spatial Processes (BioSP), INRAE, Avignon 84914, France

Correspondence: Eamonn Organ (Eamonn.Organ@ul.ie)

Abstract. Accurate estimates of wind speeds at wind turbine hub heights are crucial for both wind resource assessment and day-to-day management of electricity grids with high renewable penetration. In the absence of direct measurements, parametric models are commonly used to extrapolate wind speeds from observed heights to turbine heights. Recent literature has proposed extensions to allow for spatially or temporally varying vertical wind gradients, that is, the rate at which wind speed changes with height. However, these approaches typically assume that reference height and hub height measurements are available at the same locations, which limits their applicability in operational settings where meteorological stations and wind farms are spatially separated. In this paper, we develop a two-step spatio-temporal framework to estimate turbine height wind speeds using only open-access observations from sparse meteorological stations. First, a non-parametric generalized additive model is trained on reanalysis data to perform vertical height extrapolation. Second, a spatial Gaussian process model interpolates these hub-height estimates to wind farm locations while explicitly propagating uncertainty from the height extrapolation stage. The proposed framework enables the construction of high-resolution, sub-hourly turbine-height wind speed time series and spatial wind maps using data available in real time, capabilities not provided by existing reanalysis products. We further provide calibrated uncertainty estimates that account for both vertical extrapolation and spatial interpolation errors. The approach is validated using hub-height measurements from seven operational wind farms in Ireland, demonstrating improved accuracy relative to ERA5 reanalysis while relying solely on real-time, open-access data.

1 Introduction

Estimating wind speeds at turbine hub heights is crucial for both wind resource assessment and day-to-day management of electricity grids, and its importance increases as the share of wind power in total generation grows. In Ireland, 33% of electricity was produced by wind farms in 2025, with the Government setting a target of 80% by 2030 (Wind Energy Ireland, 2026; Department of the Environment, Climate and Communications, 2024). To meet that target, the Government has set a goal of 9 GW of onshore wind energy to be installed by 2030. There are currently approximately 300 wind farms which provide 4.8 GW of installed capacity. Based on the current average wind farm size, this would require the construction of approximately 260 new wind farms in the coming years (Sustainable Energy Authority of Ireland, 2025b).



Wind resource assessment requires knowing wind speed at turbine hub height (i.e., the height at the centre of the rotor)
25 (Murthy and Rahi, 2017), but official measurements of wind speeds are generally available only at 10 m (i.e., the standard height
of weather stations) (Oke, 2018), leading energy providers to carry out their own surveys at potential sites (Weekes and Tomlin,
2014). On-site assessments required for investors can cost from €10,000 to in excess of €100,000, so improved preliminary site
screening can lead to targeting sites with higher potential (Bailey et al., 1997; Teagasc, 2020). Such costs can be prohibitive
for small-scale or community energy projects, effectively limiting their ability to undertake rigorous resource assessments
30 and delaying or preventing otherwise viable developments. As a result, reanalysis datasets are popular products for historical
wind resource assessment, as they provide wind speeds on a regular grid, typically over several decades (Doddy Clarke et al.,
2021). However, these are often coarse in space (with grid sizes from 1 to in excess of 50 km) and time (typically on an hourly
resolution) and released with delays that make them unsuitable for real-time operational use (Ramon et al., 2019). This creates a
gap between the availability of near-surface wind observations and the need for accurate hub-height wind estimates at arbitrary
35 locations and times.

For day-to-day grid operations, accurate and real-time information on wind conditions is critical in energy systems with high
renewable penetration. In particular, errors in wind power forecasts impose substantial operational costs. In the Irish context,
Mc Garrigle and Leahy (2015) estimate that a one percentage point reduction in forecast error could yield annual savings of
approximately €2.5 million. While forecasting is not the focus of this study, accurate real-time estimates of wind speed form
40 a key input to short-term forecasting and grid management workflows. For such real-time estimation tasks, statistical methods
have been shown to improve upon pure numerical weather prediction outputs (Sweeney et al., 2020). As for wind resource
assessment, the real-time estimation for grid operations requires wind speed at turbine height.

In the absence of wind measurements at turbine hub height, a common alternative is to vertically extrapolate wind speeds
from observations at 10 m or other near-surface levels. There is extensive research on methods for extrapolating wind speeds
45 from a reference height to turbine hub height, most of which rely on simple parametric formulations, primarily the log law
and the power law (Houndekindo and Ouarda, 2025). Gualtieri (2019) reviews the classical models used in this context. These
laws generally depend on parameters relating to surface roughness or air stability. As these parameters are usually unknown or
difficult to estimate, fixed approximations are used. These approximations, however, can introduce systematic biases at some
locations and cannot capture temporal changes in the vertical wind speed gradient. To complement parametric formulations,
50 recent research has therefore focused on statistical methods that learn how wind speeds vary with height. Crippa et al. (2021)
fitted a model in which the power-law parameters vary over time and depend on weather variables other than wind speed. Other
studies have explored non-parametric approaches such as random forests and gradient-boosting models to predict wind speeds
at target heights (Yu and Vautard, 2022). These methods rely on datasets that provide wind speeds at multiple heights, such as
physical weather models or specialised remote-sensing instruments like LiDAR (Light Detection and Ranging) (Vassallo et al.,
55 2020). Model training and testing are typically carried out on multi-height datasets, meaning these studies assume that wind
speeds at several heights are co-located and that wind speeds at a reference height are available at the prediction site. This is
a strong practical limitation because in most real-world settings the location where hub-height winds are required is different



from where open-access 10 m measurements are observed, and physical weather model outputs are often too coarse to provide reliable site-specific vertical wind gradients (Gualtieri, 2022).

60 As a result, existing height-extrapolation models cannot be directly applied when hub-height winds are required at locations where only near-surface observations from distant stations are available, which is the most common operational scenario for both grid operators and early-stage site assessment. Moreover, existing approaches do not propagate height-extrapolation uncertainty into subsequent spatial prediction. Consequently, there is currently no established framework capable of estimating hub-height wind speeds at unmonitored locations using only near-surface observations from distant stations, while explicitly
65 propagating height-extrapolation uncertainty into subsequent spatial predictions.

In this context, the present paper addresses a practical but underexplored problem in wind resource modelling: estimating turbine hub-height wind speeds at locations where no on-site measurements exist. This is done using only 10-m-high wind observations recorded at distant sites. We develop a modelling framework that integrates height extrapolation and spatial interpolation within a unified probabilistic framework, allowing uncertainty from the vertical extrapolation stage to be explicitly
70 propagated into spatial prediction at new locations. This enables hub-height winds to be estimated at unobserved sites using sparse, 10 m high observations from disparate locations.

The proposed methodology consists of two linked stages: (i) non-parametric vertical extrapolation of 10 m wind observations to turbine hub heights, and (ii) spatial interpolation of these hub-height estimates to unobserved locations using Gaussian process regression. We propose performing vertical height extrapolation before spatial interpolation, arguing that wind speeds
75 are spatially smoother at hub heights. For vertical height extrapolation, we employ a non-parametric model trained on reanalysis data and then applied to meteorological stations. The resulting hub-height wind speeds are subsequently interpolated using Gaussian process (GP) regression to generate predictions at arbitrary locations. Note that although this study focuses on meteorological stations as reference wind speeds, the proposed methodology can also be applied to extrapolate turbine height wind speeds derived from numerical weather predictions or satellites (Verhoef et al., 2012), which are typically at 10 m.

80 A key contribution of the proposed framework is that it relies exclusively on open-access data streams and can therefore generate turbine-height wind estimates and spatial maps in real time, unlike reanalysis products which are released with multi-day delays. This makes the approach suitable for day-to-day management of renewable grids with high penetrations of wind energy. It can also be applied retrospectively to construct high-resolution hub-height wind-speed time series at prospective sites. Unlike reanalysis datasets, meteorological station observations are available at sub-hourly (often minute-level) resolution,
85 enabling resource assessments at a temporal resolution consistent with the nonlinear power curve of modern turbines. Recent studies have shown that higher temporal resolution wind-speed data are important for accurate wind resource and power estimation, due to the cubic relationship between wind speed and power generation (Vest and Tych, 2025). Another advantage of the framework is its ability to provide full uncertainty quantification for all predictions. This is in contrast with reanalysis data, which provide only point estimates.

90 To our knowledge, this is among the first frameworks that jointly performs height extrapolation and spatial interpolation of wind speeds while propagating uncertainty across both stages using only open-access, real-time data. This enables the



generation of prospective hub-height wind estimates and minute-level time series at arbitrary locations using only publicly available data, together with calibrated uncertainty estimates.

The layout of the paper is as follows: In Section 2 we introduce the datasets used. Section 3 briefly reviews classical vertical height extrapolation approaches, and introduces our non-parametric strategy, while also reviewing the spatial models used. Section 4 presents our results on both reanalysis data, and proprietary wind farm data. Finally, Section 5 discusses our conclusions future research directions.

2 Data

This study integrates three different data sources to develop and validate the proposed model. First, meteorological station data provide high-quality, real-time measurements of wind speed at a 10 m reference height. Second, reanalysis data, the output of physical weather models, provide wind speeds at 10 m and multiple heights above ground level, which are used to train the height extrapolation model in Section 3.2. Finally, wind speed measurements from wind farms serve as an external validation dataset to evaluate the accuracy of the complete modelling pipeline.

2.1 Weather Station Data

Ground-level wind observations are sourced from the network of 23 meteorological stations operated by Met Éireann, Ireland's national meteorological service (Met Éireann, 2025). These stations adhere to World Meteorological Organization standards (Oke, 2018) and record wind speed at a height of 10 m.

Of the 23 stations, 18 provide data at a one-minute temporal resolution, while the remaining 5 report wind speeds at an hourly resolution. To ensure temporal consistency with the wind farm dataset used for validation (Section 2.3), all station data are processed to a 10-minute resolution. For the 18 high-resolution stations, data are averaged over 10-minute intervals, while for the 5 hourly stations, a linear interpolation is applied to generate 10-minute estimates.

The geographical distribution of the stations, with the colour corresponding to median wind speeds over the year 2023, is shown in Figure 1, illustrating how coastal locations have typically higher wind speeds. Figure 2 depicts the daily cycle of average hourly wind speed at a sample of four locations. This shows a clear diurnal pattern of wind speed, a phenomenon driven by the differential heating of the Earth's atmosphere (Dai and Deser, 1999), which is consistently captured across stations.

2.2 Reanalysis Data

Reanalysis provides a gridded, historical record of the Earth's climate system by combining a physical weather model with observational data via data assimilation (Lahoz and Schneider, 2014). Running the model retrospectively with this constraint produces a more accurate and coherent dataset than the model or observations could provide alone (Lahoz and Schneider, 2014). These products therefore offer a valuable source of information on numerous weather variables but are inherently delayed, being released days or weeks after the fact, and are thus unsuitable for real-time applications (Organ et al., 2026).

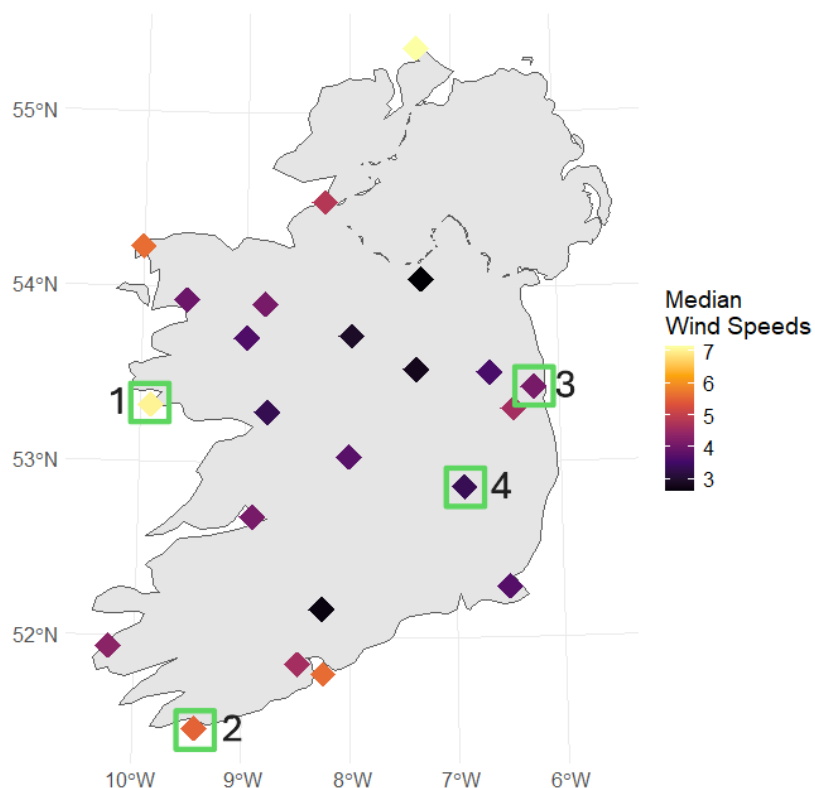


Figure 1. The location of Ireland's 23 meteorological stations. Highest wind speeds are generally seen along the west coast.

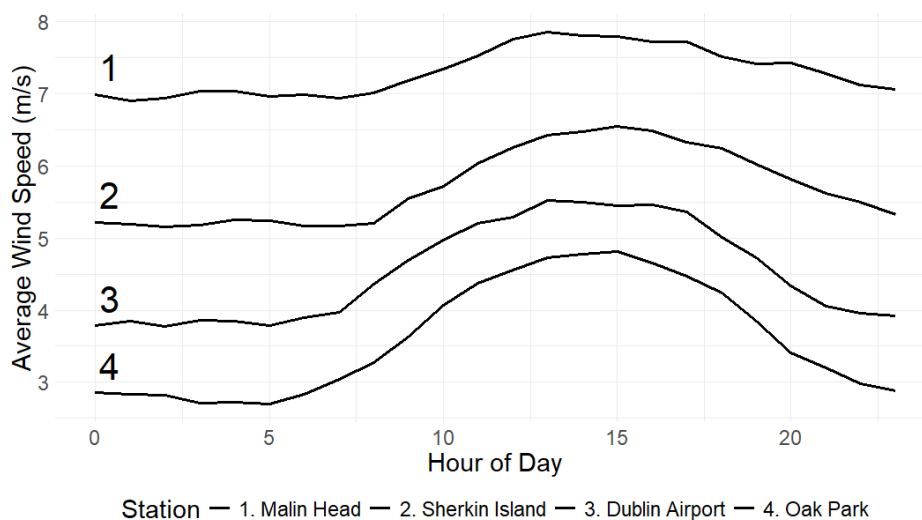


Figure 2. Comparison of hourly average speeds at four locations (highlighted in Figure 1. All stations share a similar diurnal (daily) trend).



Table 1. Summary of reanalysis and wind atlas datasets used in this study.

Dataset	Spatial resolution	Heights	Temporal information	Role in this study
ERA5	$0.25^\circ \times 0.25^\circ$ (~30 km)	Wind speeds at 10 m and 100 m	Hourly time series (1940 - present)	Benchmark for hub-height wind speed prediction
NEWA	3 km	Wind speeds at 10, 50, 75, 100 m	30-min time series (1989-2018)	Training data for height extrapolation model
GWA	250 m	Weibull parameters and long-term mean wind speeds at 10, 50 and 100 m	Mean and Weibull parameters only (no time series available)	Statistical downscaling of NEWA to station locations and mean covariate

2.2.1 ERA5

Widely used global reanalysis products for wind applications include ERA5, produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) (Hersbach et al., 2020), and NASA’s Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA2) (Gelaro et al., 2017). These datasets have become a cornerstone for renewable energy resource assessment in numerous studies (Doddy Clarke et al., 2021; Olauson, 2018). In particular, Doddy et al. (Doddy Clarke et al., 2021) identify ERA5 as the most reliable global reanalysis product for wind resource assessment over Ireland, and we therefore adopt ERA5 as a benchmark for comparison in this study. A primary limitation, however, is their relatively coarse resolution. For instance, the ERA5 data used in this study as a benchmark are provided on a $0.25^\circ \times 0.25^\circ$ grid, translating to a nominal resolution of approximately 28 km east-west and 17 km north-south at Irish latitudes.

2.2.2 New European Wind Atlas (NEWA)

To train the height extrapolation model, we require a proxy dataset containing wind speeds at both the standard reference height of 10 m and at turbine-relevant hub heights. We therefore utilise reanalysis data for this purpose. For applications requiring finer spatial scales, several high-resolution reanalysis products have been developed. In this paper, we employ the New European Wind Atlas (NEWA) as our primary data source for model training, which provides a time series of wind speeds spanning almost 30 years (Hahmann et al., 2020; Dörenkämper et al., 2020). We use the most recent full year of data available, which is 2018.

The spatial resolution of NEWA is 3 km, with wind fields available at a temporal resolution of 30 minutes. Wind speeds are provided at several heights, including 10 m, 50 m, 75 m, and 100 m. Additional heights above 100 m are also available, but these were excluded from this study as they exceed the range of turbine hub heights in our validation dataset.



Figure 3 illustrates the average wind speeds from NEWA over a full year, while Figure 4 shows wind speeds at multiple heights at four sample sites over the five-day period 1–5 January 2023. These plots show that at coastal sites, surface wind speeds (10 m) closely follow those at hub heights, whereas at inland sites, a greater disparity exists between the two. To summarise this pattern across the whole country, Figure 5 shows a map of empirical correlation between 10 m and 100 m wind speeds across the NEWA dataset.

2.2.3 Global Wind Atlas

While the NEWA provides an extensive time series of wind speeds over several years, the relative coarseness of the grid means the wind speeds at a meteorological station may not closely match the wind speed distribution of the nearest reanalysis grid cell. To address this mismatch, we use a higher-resolution dataset to statistically downscale the wind speed distribution. Specifically, the Global Wind Atlas (GWA) is used (Davis et al., 2023), which has been employed in previous studies to downscale coarser physical models (Gruber et al., 2022, 2019).

Importantly, the GWA provides only long-term climatological summary statistics rather than a time series of wind speeds, and therefore cannot be used directly for temporal modelling or height extrapolation. Instead, it is used here in complement to NEWA to impose fine-scale spatial structure on the reanalysis time series. The GWA contains summary statistics such as Weibull distribution parameters at a 250 m resolution, which better capture local topography and surface roughness effects.

Figure 3 shows average wind speeds from the GWA across Ireland. When compared to the NEWA plot, the higher spatial resolution reveals additional local variation and fine-scale spatial structure in wind speeds.

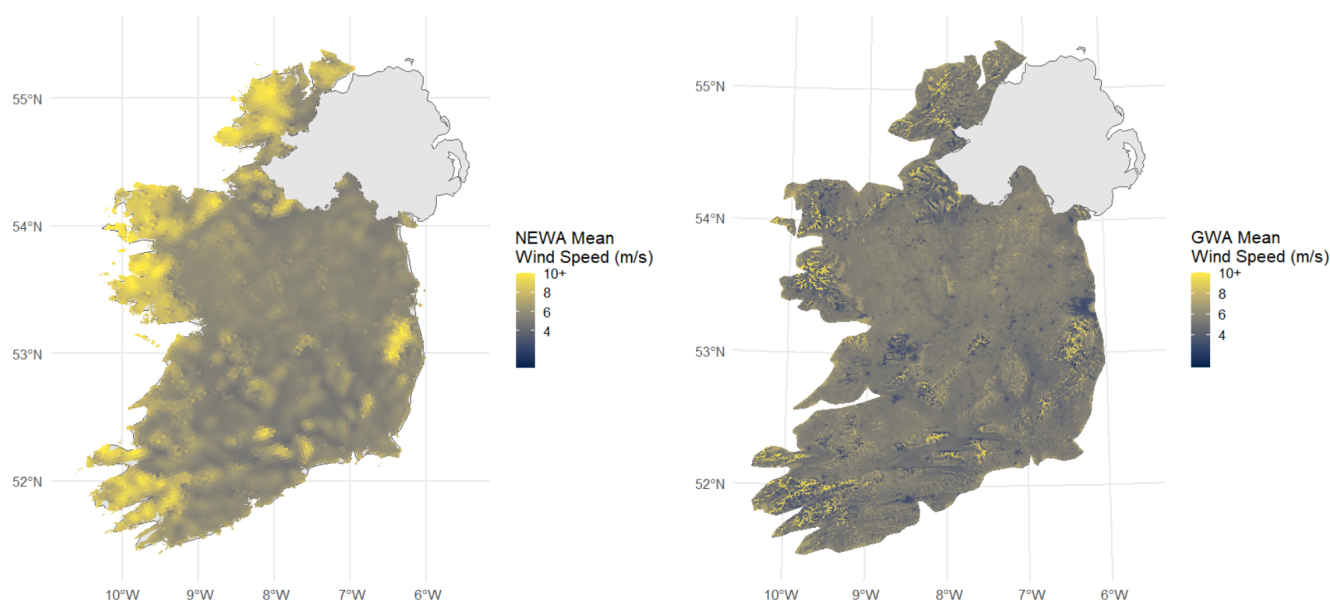
2.3 Wind Farm Data

To illustrate the potential of our framework and assess the accuracy of the proposed models, we use data from a selection of seven wind farms provided by an industry partner. Ireland currently has over 300 operational wind farms (Sustainable Energy Authority of Ireland, 2025b, a), with their locations shown in Figure 7. The counties containing the seven validation farms are highlighted in blue, though specific sites are not identified to maintain data privacy.

Each wind farm comprises of multiple turbines, each recording wind speeds independently. Therefore, model validation can be performed using either the average wind speed across the farm or the speed recorded at an individual turbine. In this study, we use two summary measures for comparison:

1. the highest recorded wind speed across all turbines at each time point, and
2. the average wind speed across the farm at each time point.

Wind speeds within a farm are influenced by the wake effect, where downstream turbines experience reduced wind speeds due to the energy extraction of upstream turbines (González-Longatt et al., 2012; Adaramola and Krogstad, 2011). We therefore interpret the maximum recorded speed as being less affected by wake losses, while the farm average reflects the aggregate effect of wake losses. Figure 6 shows the wind speeds at a single farm, comparing both the average and the speeds of each individual turbine. Since reanalysis datasets are designed to represent unobstructed, free-stream wind conditions, we expect our predicted



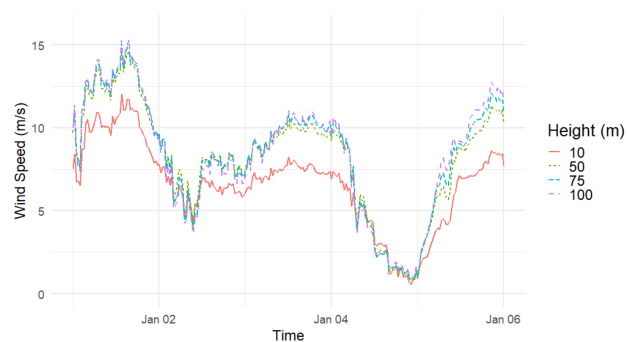
(a) New European Wind Atlas (NEWA)

(b) Global Wind Atlas (GWA)

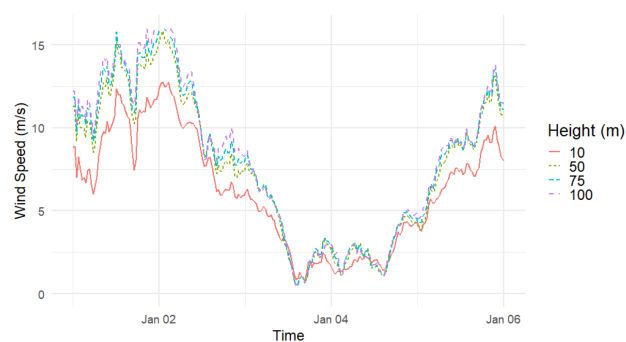
Figure 3. Comparison of mean wind speeds at 10 m for two reanalysis products. The higher resolution of the GWA (250 m vs. 3 km) resolves significantly more local topographic variation in wind speed. The NEWA is based on the most recent full year available, 2018. The GWA uses the most recent model update, release 4.0 in June 2025, which provides improved modelling and more recent land use maps (Global Wind Atlas, 2025).

wind speeds to more closely match the maximum recorded speeds, while systematically overestimating the farm average values unless wake losses are explicitly accounted for.

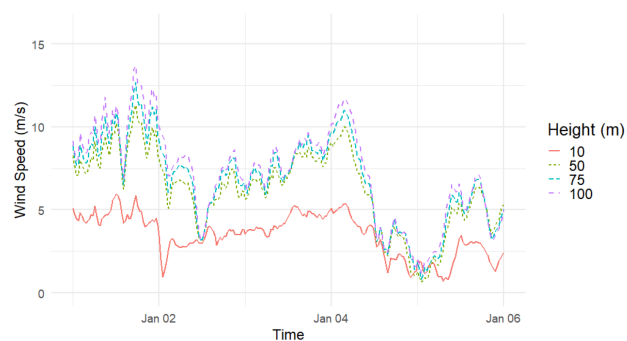
175 Turbines typically measure wind speeds at their hub, which is the height at the centre of the rotor. Hub heights can vary from 60 to 100 m for typical wind turbines found in Ireland, depending on the turbine model. Turbine wind speeds are provided at a 10 minute resolution, and all modelling in this study is therefore performed at this time scale.



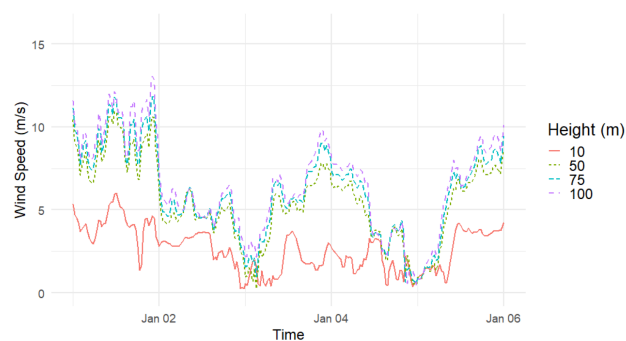
1) Malin Head



2) Sherkin Island



3) Dublin Airport



4) Oak Park

Figure 4. Wind speeds at four locations over the five-day period 1–5 January 2023 (the location of each station is highlighted in Figure 1). Each line represents wind speed at a different height (10 m, 50 m, 75 m, 100 m). We choose a sample of four locations that cover both coastal and inland sites.

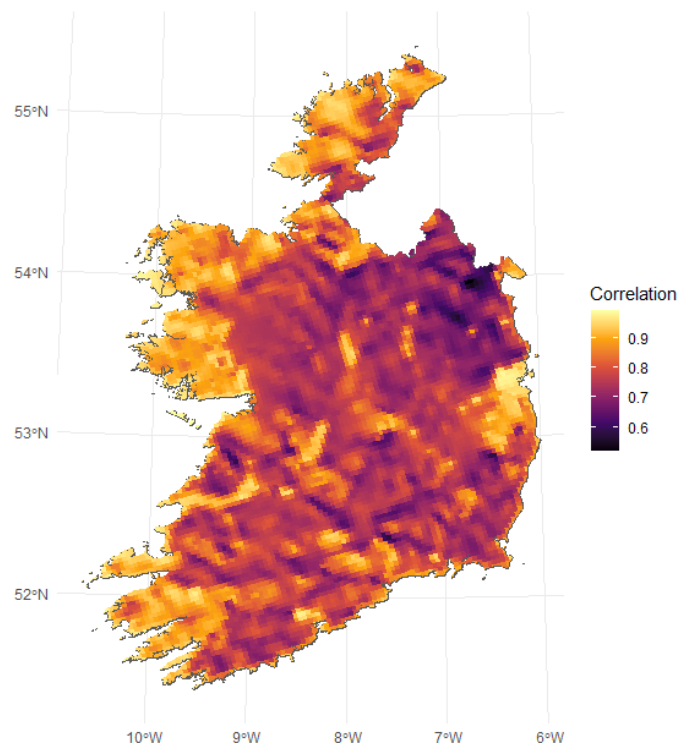


Figure 5. Empirical Pearson correlations between wind speeds at 10 m and 100 m. Correlations are generally highest in coastal and upland locations, areas with high surface wind speeds.

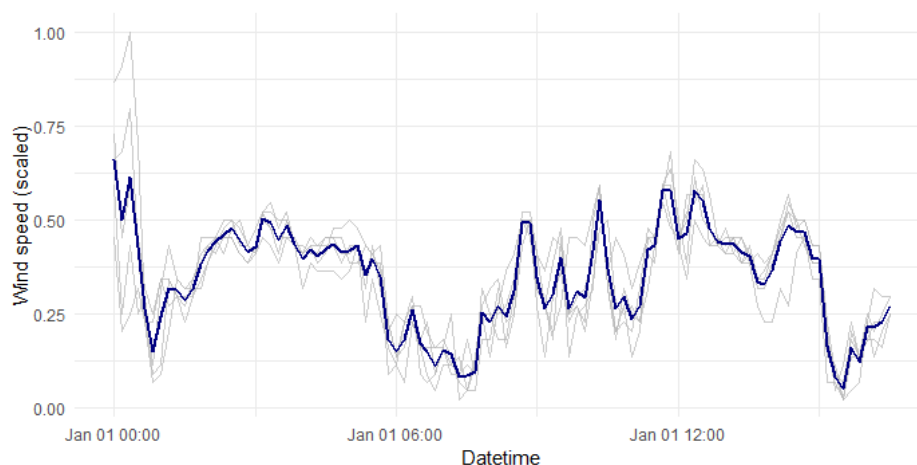


Figure 6. Individual turbine scaled wind speeds (grey) and average wind speed (navy) at a single wind farm over a five-day period.

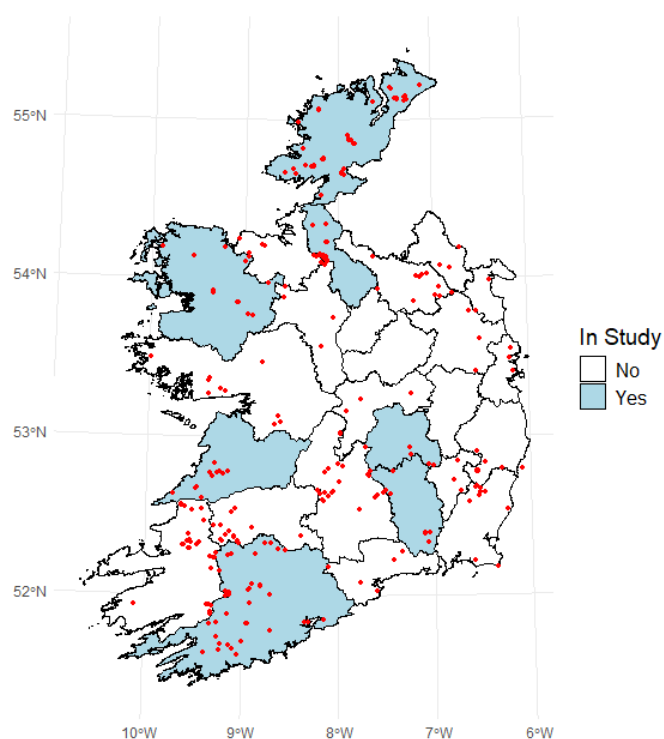


Figure 7. The locations of Ireland’s wind farms. There are over 300 wind farms in Ireland. The names and locations of wind farms are publicly available and published by the Sustainable Energy Authority of Ireland (Sustainable Energy Authority of Ireland, 2025b, a). The highlighted counties contain the seven wind farms used for validation.



3 Modelling Framework

The objective of our modelling framework is to use weather station observations to predict wind speeds at wind farm locations and turbine hub heights. This introduces two challenges: (i) a spatial interpolation process is required to interpolate wind speeds from meteorological stations to wind farm sites, and (ii) a vertical extrapolation model is needed to predict wind speeds at turbine hub height, conditional on 10m observations, which are the standard reporting height for meteorological stations.

A practical constraint is that wind speed observations at turbine hub heights are typically proprietary and not publicly accessible. Consequently, stakeholders often have limited or no access to this data. In this study, wind farm hub height measurements are therefore used only as an external test set, while all model training relies exclusively on the publicly available datasets described in Section 2.

Our prediction framework addresses these challenges through a sequential two-step approach, illustrated in Figure 8:

1. **Vertical height extrapolation:** A generalised additive model (GAM) is trained on reanalysis data to learn vertical wind-speed profiles and to predict turbine hub-height wind speeds from 10 m reference observations at meteorological stations.
2. **Spatial interpolation:** The resulting hub-height wind speed predictions are then interpolated to target wind farm locations using a spatial statistical model

The order of these steps is motivated by the fundamental meteorological principle that wind fields are smoother and exhibit longer-range spatial correlation at higher altitudes, free from the fine-scale turbulence induced by surface roughness (Schlez and Infield, 1998). As illustrated in Figure 9, the mean wind field at 100 m is considerably smoother than at 10 m, where local topography and vegetation introduce strong small-scale variability. Performing height extrapolation first leverages this smoother, more spatially coherent field for the subsequent interpolation step, which is likely to improve accuracy.

3.1 Reanalysis Wind Speeds Preprocessing

Before fitting the height extrapolation models, we preprocess the reanalysis data to address its inherent spatial coarseness. While the New European Wind Atlas (NEWA) provides a high-quality temporal record, its 3 km resolution may not fully capture the local wind climate at a specific point, such as a meteorological station location. To incorporate finer-scale topographic effects, we statistically downscale the NEWA data using the high-resolution wind climatology from the Global Wind Atlas (GWA).

The downscaling procedure is a quantile-quantile transformation at each of the NEWA height levels (Michelangeli et al., 2009), which maps the empirical distribution of the NEWA data at a grid point to a theoretical distribution defined by the GWA's parameters at the location of the reference sites. This process effectively imparts the high-resolution spatial structure of the GWA's climatology onto the NEWA time series.

Let $\mathbf{w}_{\text{NEWA}} = (w_1, \dots, w_n)$ be a time series of wind speeds of size n from the nearest NEWA grid point to a meteorological station, with empirical cumulative distribution function (CDF) $F_n(w) = \frac{\#(w_i \leq w)}{n}$. The empirical cumulative probability (or percentile rank) associated with an observation w_i is $p_i = F_n(w_i)$.

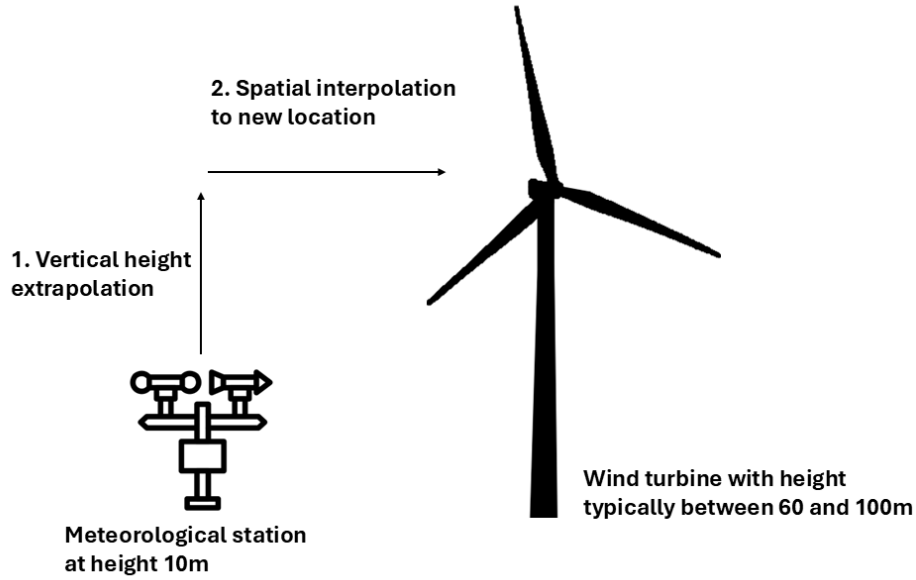


Figure 8. Schematic of the two-step process: (1) vertical extrapolation from 10 m to turbine hub height, followed by (2) spatial interpolation from meteorological stations to wind farm locations.

210 We assume that the long-term wind speed distribution at a location s follows a parametric distribution with CDF $\hat{F}_s(w; \Theta_s)$, where Θ_s denotes the distribution parameters. For this purpose, we adopt the two-parameter Weibull distribution, which is widely used to model wind speed climatology (Carta et al., 2009; Wais, 2017; Bowden et al., 1983). Its probability density function (PDF) and CDF are given by:

$$f(x; k, \lambda) = \frac{k}{\lambda} \left(\frac{x}{\lambda} \right)^{k-1} \exp \left[- \left(\frac{x}{\lambda} \right)^k \right], \quad F(x; k, \lambda) = 1 - \exp \left[- \left(\frac{x}{\lambda} \right)^k \right], \quad (1)$$

215 where $k > 0$ and $\lambda > 0$ are the shape and scale parameters, respectively. These parameters are directly available from the Global Wind Atlas (GWA) at 250 m resolution (Davis et al., 2023), and the Weibull distribution has been shown to accurately represent Irish wind speed distributions (Organ et al., 2026). For each station location, we extract the corresponding GWA parameters (k_s, λ_s) , define $\hat{F}_s$ using Equation (1), and apply the quantile mapping in Equation (2) to the NEWA time series to obtain the downscaled dataset.

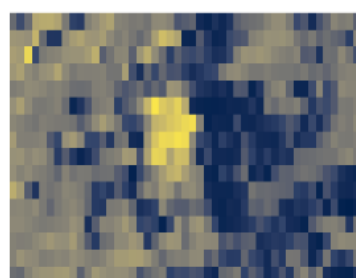
220 The downscaled wind speed, $\tilde{w}_{s,i}$, is obtained by applying the inverse CDF transform using the GWA-derived theoretical distribution:

$$\tilde{w}_{s,i} = \hat{F}_s^{-1}(p_i) = \hat{F}_s^{-1}(F_n(w_i)). \quad (2)$$

As the downscaled NEWA dataset only reports wind speeds at one reference height and three target heights (50 m, 75 m, 100 m), there is a risk that models trained directly on these values may overfit to the discrete levels or fail to capture the



(a) Satellite image of the region
 Map Data: © 2026 Google Earth Pro, Imagery: © 2026 Airbus



10m GWA Mean
 Wind Speed (m/s)

10+
 8
 6
 4
 2



100m GWA Mean
 Wind Speed (m/s)

10+
 8
 6
 4
 2

(b) GWA mean wind speed at 10 m

(c) GWA mean wind speed at 100 m

Figure 9. Comparison of wind speeds at a sample wind farm (outside the validation dataset) in the west of Ireland: (a) satellite image. (b) GWA mean wind speed at 10 m showing fine-scale variability. The average speeds within the image have a range of 9.7 m/s and a standard deviation of 2.1 m/s. (c) GWA mean wind speed at 100 m showing smoother wind field at turbine height. The average wind speeds have a range of 3.4 m/s and a standard deviation of 0.65 m/s. The large amounts of forestry around the turbines cause large variations in average wind speeds at 10 m, while the 100 m wind speeds significantly lower levels of variation.

continuous effect of height on wind speed. To address this, we construct a smooth vertical profile of wind speed at each time and location by interpolating the three target heights with a quadratic curve. From this fitted curve, wind speeds are then sampled at 5 m intervals, producing a denser set of pseudo-observations. A 5 m spacing was chosen as a practical balance: it provides sufficient vertical resolution to train a non-parametric extrapolation model while avoiding unnecessary computational overhead.

3.2 Height Extrapolation

Wind speed extrapolation from observations at a reference height (typically 10 m) to turbine hub height is traditionally performed using parametric models such as the power law or logarithmic law. A comprehensive overview is given in Gualtieri (2019), which found the power log to be the most popular and generally most reliable. In its generic form, the power law is



defined as

$$W_h = W_{10} \left(\frac{h}{10} \right)^\alpha,$$

where W_h and W_{10} are wind speeds at heights h and 10 m, respectively, and α is a wind shear parameter dependent on surface roughness and atmospheric stability. Although α is known to vary in space and time, fixed approximations are often used for simplicity, with $\alpha = 1/7$ being the most common choice (Houndekindo and Ouarda, 2025; Panofsky and Dutton, 1984).

Extensions have been proposed to account for this variability by allowing α to depend on covariates. For instance, Crippa et al. (2021) modelled $\alpha(t)$ as a function of diurnal harmonics, with a separate model fit to each location in their dataset:

$$\alpha(t) = \alpha_0 + \sum_{i=1}^P \beta_i \sin\left(\frac{2\pi t i}{24}\right) + \beta'_i \cos\left(\frac{2\pi t i}{24}\right), \quad (3)$$

where α_0 is a constant mean term, and β_i and β'_i are the coefficients for P periodic harmonics. Such approaches substantially improved predictions relative to fixed α values. However, our analysis of reanalysis data indicates that empirically derived α values display high variability and strong, non-linear interactions with surface wind speed (Figure 10), limiting the effectiveness of parametric approaches that rely on a simplified structural form. The high variability suggests that direct non-parametric modelling of wind speeds may be preferable.

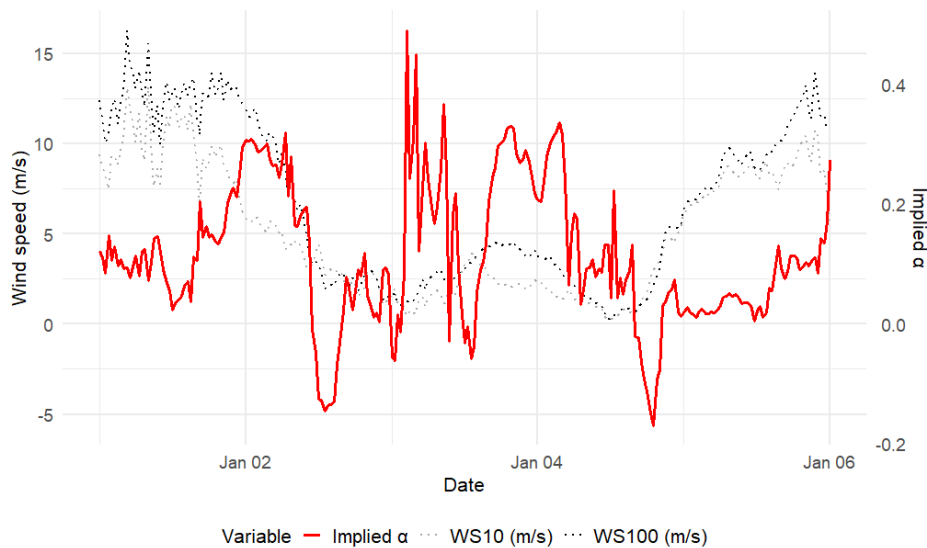


Figure 10. Five days of 10 m and 100 m wind speeds at Malin Head, labelled (1) in Figure 4. The α value at each time point derived from the two wind measurements is plotted. This plot illustrates that α shows a high amount of variability, and constant or periodic harmonics may struggle to fully capture the relationship

Previous studies using non-parametric approaches also provide good accuracy for wind height extrapolation predictions (Yu and Vautard, 2022; Vassallo et al., 2020). We choose to adopt a flexible, non-parametric approach using a Generalized Additive



Model (GAM) (Wood, 2017). GAMs have been shown to achieve accuracy comparable to machine learning methods, but offer parsimonious inference and interpretable outputs (Doohan et al., 2026). In our framework the GAM predicts the *square root* of hub-height wind speeds in the processed NEWA dataset. This transformation is commonly used to reduce positive skewness and to approximate Gaussianity in wind-speed data (Lenzi and Genton, 2020). As described in Section 3.1, the NEWA time series are first statistically downscaled using Weibull distributions derived from the Global Wind Atlas in order to encode fine scale spatial structure. For Irish wind climates, the Weibull shape parameter is typically close to 2 (Organ et al., 2026). If a wind speed follows a Weibull distribution with shape parameter k and scale parameter λ , then its square root also follows a Weibull distribution with shape parameter $2k$ and scale parameter $\sqrt{\lambda}$. Consequently, the square-root transformation yields a distribution with shape parameter close to 4, a regime in which the Weibull distribution exhibits low skewness and is well approximated by a Gaussian distribution (Cui and Xie, 2003). This motivates the use of a Gaussian likelihood on the square-root-transformed scale, which simplifies model fitting while remaining consistent with the empirical distribution of Irish wind speeds.

Predictors include the square root of 10 m wind speed, turbine hub height, diurnal cycles, and wind direction components. These covariates are chosen as crucially, they are all available in real time at meteorological stations. The model is specified as:

$$\sqrt{W_{h,i}} = \beta_0 + s_1(\sqrt{W_{10,i}}) + s_2(h_i) + s_3(h_i, W_{10,i}) + \beta_1 \sin\left(\frac{2\pi t_i}{24}\right) + \beta_2 \cos\left(\frac{2\pi t_i}{24}\right) + \beta_3 U_i + \beta_4 V_i + \varepsilon_i \quad (4)$$

where β_0 is an intercept term, and $\beta_1, \dots, \beta_4$ are coefficients associated with the linear covariates. Here, $i = 1, \dots, n$ indexes independent observations for a given meteorological station, each corresponding to a specific time point and height level. $W_{h,i}$ and $W_{10,i}$ denote wind speeds at height h_i and 10 m for observation i . t_i denotes the hour of the day associated with observation i . U_i and V_i are the east–west and north–south components of the unit wind direction vector at 10 m. All smooth terms use the default spline bases in `mgcv` (thin-plate regression splines), with smoothing parameters estimated via restricted maximum likelihood (REML). Basis dimensions were checked using standard `mgcv` diagnostics. s_2 and s_3 were left at their default values, while for s_1 the basis dimension was increased until the k-index was close to 1 (Wood, 2017). The error term ε_i represents independent residual variation and is assumed to follow $\varepsilon_i \sim \mathcal{N}(0, \sigma_s^2)$, where σ_s^2 is the station-specific residual variance for the model fitted at meteorological station s .

Figure 11 contains a sample of training curves as well as model fits to show how the predicted wind speeds vary with reference height wind speed and target height.

A separate model is fit for each meteorological station using reanalysis wind speeds extracted from the nearest NEWA grid point and statistically downscaled using the Global Wind Atlas, as described in Section 3.1, accounting for local effects such as surface roughness. Each training observation corresponds to a specific time point and height and are treated as independent observations. While this formulation ignores temporal and vertical dependence in the reanalysis data, it is consistent with the intended operational use of the model, where predictions at any given time must rely solely on the 10 m wind observation available at that location. This setup ensures that model training and evaluation reflect the information constraints present in

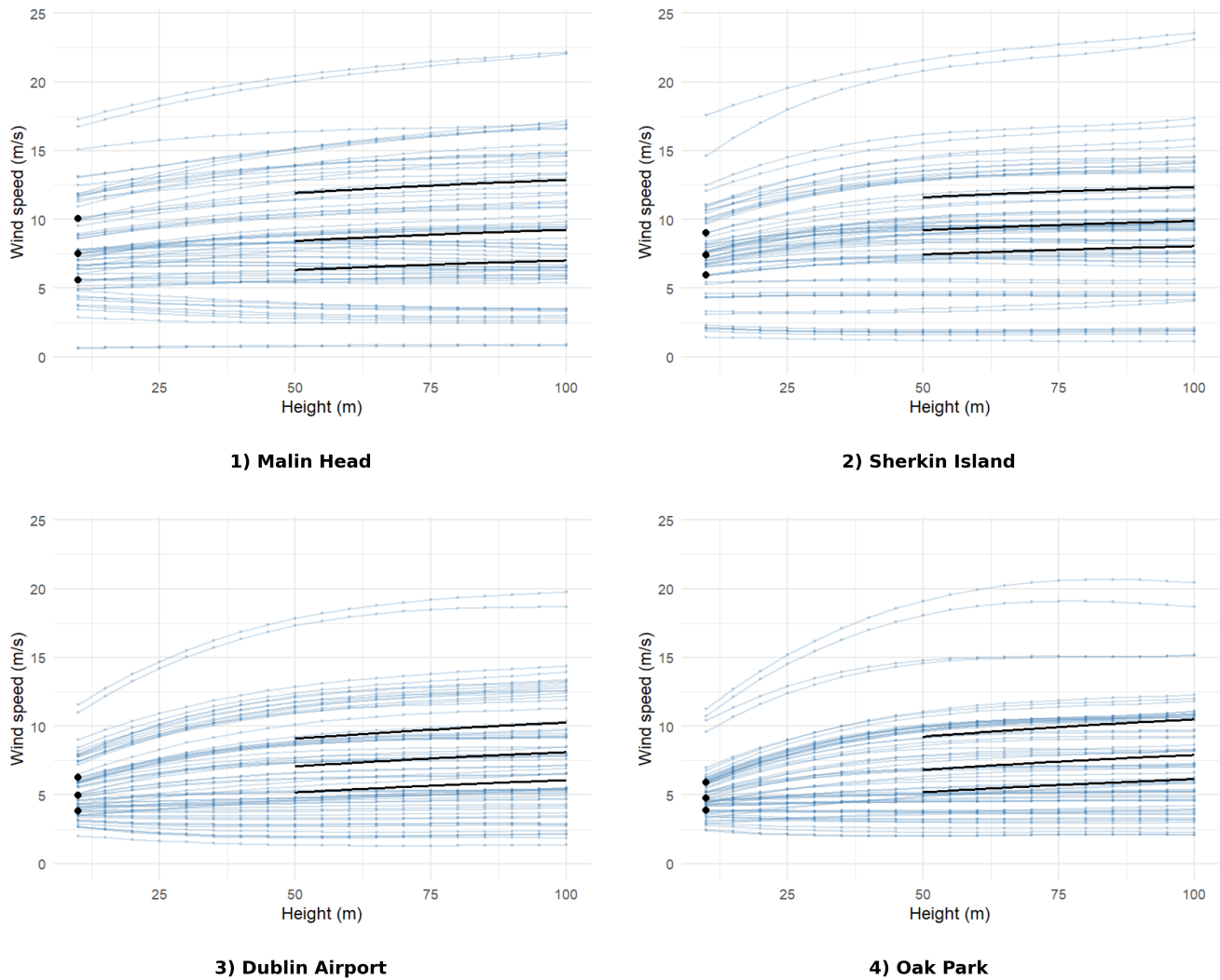


Figure 11. Representative vertical wind-speed profiles from the processed reanalysis training data at four meteorological stations. Light blue curves show interpolated vertical profiles at 13:00 across sample days, while black curves show fitted GAM predictions evaluated at representative 10 m wind-speed regimes over turbine-relevant heights (50–100 m) (25th, 50th and 75th percentiles). The figure illustrates height dependence and variation in the vertical gradient with reference wind speed.



real-time applications. Implementation uses the `mgcv` package in R, with efficient estimation (typically under 10 seconds per model on a standard laptop). The process is easily parallelised if fitting is required for many locations.

Model uncertainty is quantified directly from the fitted GAMs. For each station specific GAM, residual variance is estimated under the assumption of Gaussian errors with constant variance on the square root scale. This residual variance is used as a measure of uncertainty in the height-extrapolated wind speeds at each station. Figure 12 shows the resulting station-specific variance estimates.

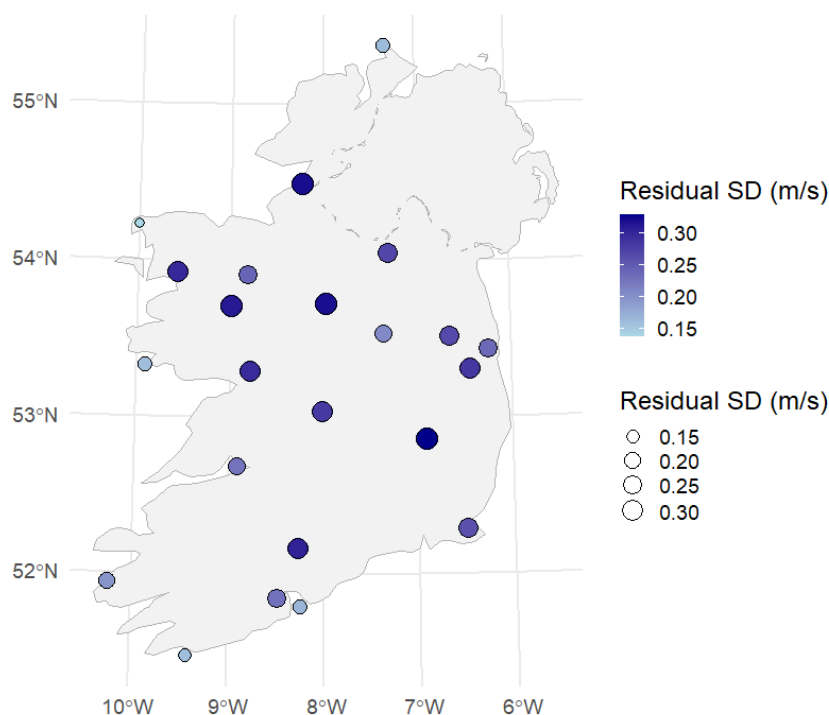


Figure 12. Estimated standard deviation σ_s of the height extrapolation model at each wind farm location.

Uncertainty is generally higher at inland sites, while coastal stations—despite higher mean wind speeds—exhibit lower extrapolation uncertainty, possibly due to smoother vertical wind profiles near open seas. As seen in Figure 4 and Figure 5, at coastal location, wind speeds at 10 m more closely follow the pattern shown at turbine height.

3.3 Spatial Interpolation

To interpolate wind speeds from observed to unobserved locations, we model wind speed as a Gaussian Process (GP) (Banerjee et al., 2003; Williams and Rasmussen, 2006). Spatial interpolation is performed at a fixed turbine hub height, as wind speeds at greater heights are generally smoother and less affected by surface obstacles. Following Section 3.2, spatial inference is



295 performed using the square root of wind speeds, which is approximately Gaussian. For notational clarity, the dependence on height is suppressed throughout this section.

The mean structure of the spatial process incorporates information from the Global Wind Atlas (GWA), with the GWA mean square-root wind speed at location s and at the target height h , denoted μ_s^h , used as a covariate. The mean square-root wind speed is derived from the GWA Weibull parameters (Organ et al., 2026). As the GWA is available only at fixed reference
 300 heights, which generally differ from turbine hub heights, a power-law relationship is used to interpolate the mean wind speed to the target height h . Let μ_{GWA} denote the vector of GWA-derived mean square-root wind speeds at the observation locations. The mean structure is specified as

$$\boldsymbol{\mu} = \beta_0 + \beta_1 \boldsymbol{\mu}_{\text{GWA}},$$

where β_0 and β_1 are regression coefficients. This provides a spatially varying baseline that reflects long-term climatological
 305 differences between locations. While other spatial covariates such as elevation and distance to the coast have been used in wind speed studies, the GWA mean implicitly incorporates these effects through the physical modelling used to generate the atlas.

The dataset comprises Met Éireann wind speed observations recorded at 10 minute intervals over the full year 2023 (52,560 time points). Fitting a fully spatio-temporal Gaussian process (GP) model to data of this size would be computationally demanding, as exact GP inference scales as $\mathcal{O}(n^3)$. While scalable approximations such as composite likelihood or Vecchia-type
 310 methods could be considered, incorporating an explicit temporal dependence structure is unlikely to substantially improve spatial interpolation in this setting, since observations are available at the same fixed locations for all time points.

We therefore assume temporal independence and treat each time point as an independent realization of the spatial process, using temporal replication to improve estimation of the spatial covariance structure. This approach substantially reduces computational cost while retaining the information needed for accurate spatial interpolation at each time step. To allow for
 315 seasonal variation in both the mean structure and spatial covariance, a separate spatial model is fitted for each calendar month, using the full 10 minute time series within that month for inference and prediction. This monthly modelling strategy reduces computational cost and enables inference to be performed on standard hardware.

The hub-height wind speeds used as inputs to the spatial model are not direct observations but estimates obtained from the height-extrapolation step in Section 3.2. To account for this additional source of uncertainty, we treat the square-root-
 320 transformed wind speeds as noisy observations of an underlying latent Gaussian process, with an observation-error variance informed by the height-extrapolation model. As discussed in Section 3.2, the square root transform provides a suitable Gaussian approximation for wind speeds, which allows for efficient spatial inference.

Let $\mathbf{W}_t$ denote the vector of square-root-transformed wind speeds observed at meteorological station locations at time t , for $t = 1, \dots, n_t$. The hub-height wind speeds used as inputs to the spatial model are not direct observations but estimates obtained
 325 from the height-extrapolation step in Section 3.2. To account for this additional source of uncertainty, we treat $\mathbf{W}_t$ as noisy observations of an underlying latent spatial process.

$$\sqrt{\mathbf{W}_t} \sim MVN(\mathbf{f}_t, \boldsymbol{\Sigma}_{\text{GAM}} + \sigma_\epsilon^2 \mathbf{I}),$$



where $\mathbf{f}$ is the latent spatial process evaluated at the observation locations, Σ_{GAM} is a diagonal matrix containing the station-specific variance estimates $\sigma_{s_i}^2$ from the height-extrapolation model at locations s_i , and $\sigma_\epsilon^2 \mathbf{I}$ is a nugget term capturing unresolved microscale variability.

$$\mathbf{f}_t \sim GP(\boldsymbol{\mu}, k(\cdot, \cdot))$$

where $\boldsymbol{\mu}$ is the mean defined above

We assume independent measurement errors at each location, making $\Sigma_{\text{GAM}} + \sigma_\epsilon^2 \mathbf{I}$ diagonal. Unlike the standard homoscedastic assumption (equal variance at all sites), we allow for heteroscedasticity by setting each diagonal entry of Σ_{GAM} to $\sigma_{s_i}^2$, the estimated variance of the residuals from the GAM predictions at location s_i . This provides a location-specific measure of uncertainty. As reanalysis data are often over-smoothed, an additional common nugget term $\sigma_\epsilon^2 \mathbf{I}$ is added to the diagonal to capture unresolved microscale variability, which is learned from the data.

Assuming independence between the latent process and measurement error, the marginal distribution of the observed data for each becomes:

$$\sqrt{\mathbf{W}_t} \sim MVN(\boldsymbol{\mu}^h, \Sigma_f + \Sigma_{\text{GAM}} + \sigma_\epsilon^2 \mathbf{I}).$$

The latent GP employs a Matérn covariance function, which is widely used for spatial processes (Banerjee et al., 2003). To simplify inference and avoid identifiability issues between range and smoothness parameters, we fix the smoothness parameter $\nu = 1$, yielding:

$$\Sigma_{f,ij} = \text{cov}(s_i, s_j) = \kappa d K_1(\kappa d),$$

where d denotes the Euclidean distance between sites s_i and s_j , and K_ν is the modified Bessel function of the second kind with parameter ν . κ is a scale parameter, which smaller values of κ denote a longer spatial correlation.

Assuming independence across time points, the log-likelihood is given by

$$\log L(\Theta) \propto -\frac{1}{2} \sum_{t=1}^{n_t} \left[(\sqrt{\mathbf{W}_t} - \boldsymbol{\mu})^\top \mathbf{C}^{-1} (\sqrt{\mathbf{W}_t} - \boldsymbol{\mu}) + \log \det(\mathbf{C}) \right], \quad (5)$$

where $\mathbf{C} = \Sigma_f + \Sigma_{\text{GAM}} + \sigma_\epsilon^2 \mathbf{I}$ and $\Theta = \{\kappa, \sigma_f, \sigma_\epsilon, \beta_0, \beta_1\}$ denotes the set of model parameters. n_t denotes the number of unique 10 minute points in a given month (e.g., $n_t = 4464$ for a 31 day month). The optimization is carried out in R using the *optim* function with the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm.

As multivariate optimisation routines can be sensitive to starting values, initial values for the spatial range and variance parameters were chosen based on estimates reported in previous studies of Irish wind speeds (Organ et al., 2026). Fixed effect coefficients were initialised using ordinary least squares estimates.

Once optimal hyperparameters are obtained from the BFGS algorithm, the posterior mean and variance of the wind speed at wind farm locations can be derived. Let $\sqrt{\mathbf{W}}$ denote the vector of observed wind speeds at station locations S , and $\sqrt{\mathbf{W}^*}$ the



(unobserved) vector of wind speeds at wind farm locations WF . We define the joint covariance matrix of all locations as

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{S,S} & \mathbf{C}_{S,WF} \\ \mathbf{C}_{WF,S} & \mathbf{C}_{WF,WF} \end{bmatrix},$$

where each block $\mathbf{C}$ denotes the covariance between locations in sets S and WF .

360 Then, the conditional (posterior) distribution of $\sqrt{W^*}$ given $\sqrt{W}$ and parameters Θ is given by standard GP results (Rue and Held, 2005):

$$\sqrt{W^*} \mid \sqrt{W}, \Theta \sim \mathcal{N}\left(\mu^* + \mathbf{C}_{WF,S} \mathbf{C}_{S,S}^{-1} (\sqrt{W} - \mu), \mathbf{C}_{WF,WF} - \mathbf{C}_{WF,S} \mathbf{C}_{S,S}^{-1} \mathbf{C}_{S,WF}\right).$$

This posterior distribution provides both predicted hub-height wind speeds and their associated uncertainty at each wind farm location.

365 Taken together, the preprocessing, height extrapolation, and spatial interpolation form an end-to-end framework for predicting turbine hub height wind speeds at unobserved locations. Once trained, the framework can be applied in real time using only 10 m wind speed and direction observations from meteorological stations. These inputs are first extrapolated to hub height and then spatially interpolated to target locations, producing turbine-height wind speed predictions without requiring access to turbine measurements or reanalysis data.

370 4 Results

The results are presented in two parts. First, we validate the height extrapolation model (Section 3.2) using a held-out test set from the reanalysis data. Second, we evaluate the full two-step framework—extrapolation followed by spatial interpolation (Section 3.3)—against the external validation dataset of hub-height measurements from operational wind farms.

4.1 Accuracy on reanalysis data

375 We first benchmark the performance of our height extrapolation model against parametric baselines using the NEWA reanalysis dataset. The first baseline employs the power law with a constant shear exponent $\alpha = 1/7$, a common approximation for neutral atmospheric conditions over smooth terrain (Panofsky and Dutton, 1984). The second baseline is a simplified, temporally varying α model following Crippa et al. (2021) (Section 3.2), which incorporates diurnal harmonics but omits the temperature gradient to rely only on predictors available in real-time. These are compared against the GAM proposed in Section 3.2.

380 A separate height-extrapolation model is fitted for each meteorological station using reanalysis wind speeds extracted from the nearest NEWA grid point and statistically downscaled using the Global Wind Atlas, as described in Section 3.1. Predictions at any given time must rely solely on the 10 m wind observation available at that location as models inputs. This setup ensures that model evaluation reflects the information constraints present in real-time applications.

All models are trained on 80% of the NEWA reanalysis data and evaluated on the remaining 20% holdout set. The root mean square error (RMSE) for each approach is presented in Table 2.



Model	RMSE (m/s)
Power Law ($\alpha = 1/7$)	2.92
Temporally Varying α	1.53
GAM (Proposed)	1.25

Table 2. Comparison of height extrapolation accuracy on the reanalysis test set.

The results demonstrate that both temporally varying α and our proposed GAM offer substantial improvements over the constant power law. The proposed GAM provides a further, meaningful increase in accuracy, likely due to its ability to capture non-linear relationships between the 10 m wind speed and the target height, as well as its incorporation of additional covariates like wind direction.

390 4.2 Accuracy at wind farms

We evaluate the end-to-end performance of our two-step framework using an external test set comprising hub-height wind speed measurements from seven operational wind farms (Section 2.3). Using one year of 10-minute data, predictions are generated exclusively from publicly available datasets and the proposed modelling framework. For comparison, we also generate predictions from the ERA5 reanalysis dataset, a commonly used benchmark in renewable energy applications (Olauson, 2018; 395 Doddy Clarke et al., 2021).

ERA5 is provided at hourly temporal resolution, whereas both the wind farm observations and the proposed framework operate at a 10-minute resolution. To enable a comparison at this finer time scale, wind speeds at 10 m and 100 m were extracted from the nearest ERA5 grid point to each wind farm location. The resulting hourly time series were then linearly interpolated to a 10-minute resolution to align with the wind farm data. This interpolation does not introduce additional physical 400 information, but represents a simple baseline approach for deriving sub-hourly wind speeds from an hourly reanalysis product. A site-specific power-law relationship was subsequently applied at each time point to estimate wind speeds at the actual turbine hub height.

Predicted wind speeds are evaluated against two validation observations defined in Section 2.3: the maximum observed wind speed (a proxy for the unobstructed free-stream wind) and the farm average wind speed (which incorporates wake effects). As 405 our model predicts the free-stream wind, its predictions are systematically higher than the farm average. To account for this and to facilitate comparisons, we apply simple empirical wake loss adjustments to predicted wind speed of 10%, 15%, and 20%, reflecting typical values reported in the literature (González-Longatt et al., 2012; Pryor and Barthelmie, 2024; Adaramola and Krogstad, 2011).

The results are summarised in Tables 3 and 5, which evaluate model performance against two validation metrics. 410 Table 3 compares predicted wind speeds against the maximum observed wind speed at each farm, which serves as a proxy for the unobstructed wind speed. Under this metric, the proposed model achieves both a lower root mean squared error (RMSE) and a substantially smaller mean residual (bias) than the ERA5 benchmark, indicating improved accuracy and reduced system-



atic error in estimating unobstructed hub-height wind speeds. This suggests that combining data-driven height extrapolation with spatial interpolation better captures local wind regimes than relying on coarse-grid reanalysis products alone, even when
 415 evaluated at hub height.

To assess whether this improvement is consistent across locations, Table 4 reports model performance at each individual wind farm. The proposed model achieves lower RMSE than ERA5 at three out of seven wind farms, with particularly large improvements observed at WF3 and WF6. While ERA5 performs well at several individual sites, its errors are notably larger at others, resulting in higher overall variability in performance across locations. In contrast, the proposed framework provides
 420 more stable predictions across sites, particularly in terms of bias, leading to improved aggregated accuracy.

While ERA5 exhibits marginally higher correlation at most individual sites, pooled correlation values across all wind farms are similar for both models. This arises because the pooled correlation reflects both temporal variability within sites and spatial variability between sites. The proposed framework improves representation of site-to-site variability, which offsets its slightly weaker within-site temporal correlation when evaluated across all locations.

Table 5 compares predictions against the farm-average wind speed, which reflects the combined effects of turbine wakes and local flow interactions. As the proposed framework predicts unobstructed wind speeds, the unadjusted model overestimates wind speeds and exhibits elevated RMSE when evaluated against farm average observations. However, after applying simple empirical wake-loss adjustments, the agreement improves substantially. In particular, the 15% wake-loss adjustment yields the lowest RMSE and a mean residual close to zero, indicating stronger agreement with observed farm average wind speeds.
 425
 430 Correlation coefficients are unchanged by the wake loss adjustment, as the adjustment corresponds to a linear scaling of the predicted wind speeds, which does not affect correlation. The sensitivity analysis across multiple wake-loss values demonstrates that modest wake corrections are sufficient to align raw predictions with farm level averages.

Model	RMSE (m/s)	Mean Bias (m/s)	Correlation
Proposed Model	1.90	-0.50	0.86
ERA5	2.14	1.18	0.86

Table 3. Prediction accuracy aggregated across all seven wind farms, evaluated against the **maximum** observed wind speed at each farm (proxy for unobstructed free-stream wind speeds). Mean bias is defined as observed minus predicted, so positive values indicate underprediction and negative values indicate overprediction.

While predictive accuracy is comparable to or better than ERA5 under both validation metrics, the proposed framework additionally provides full predictive uncertainty, combining uncertainty from both height extrapolation and spatial interpolation.

Figure 13 visualizes predicted hub-height wind speeds and 95% prediction intervals at four sample wind farms over a five day period. The prediction closely follows the observed temporal trends, as well as rapid wind ramps. Observed wind speeds generally fall within the predicted uncertainty bounds, indicating that uncertainty estimates are well calibrated. Differences between the farm average and maximum observed wind speeds are also illustrated, consistent with the presence of wake effects within the farms.
 435

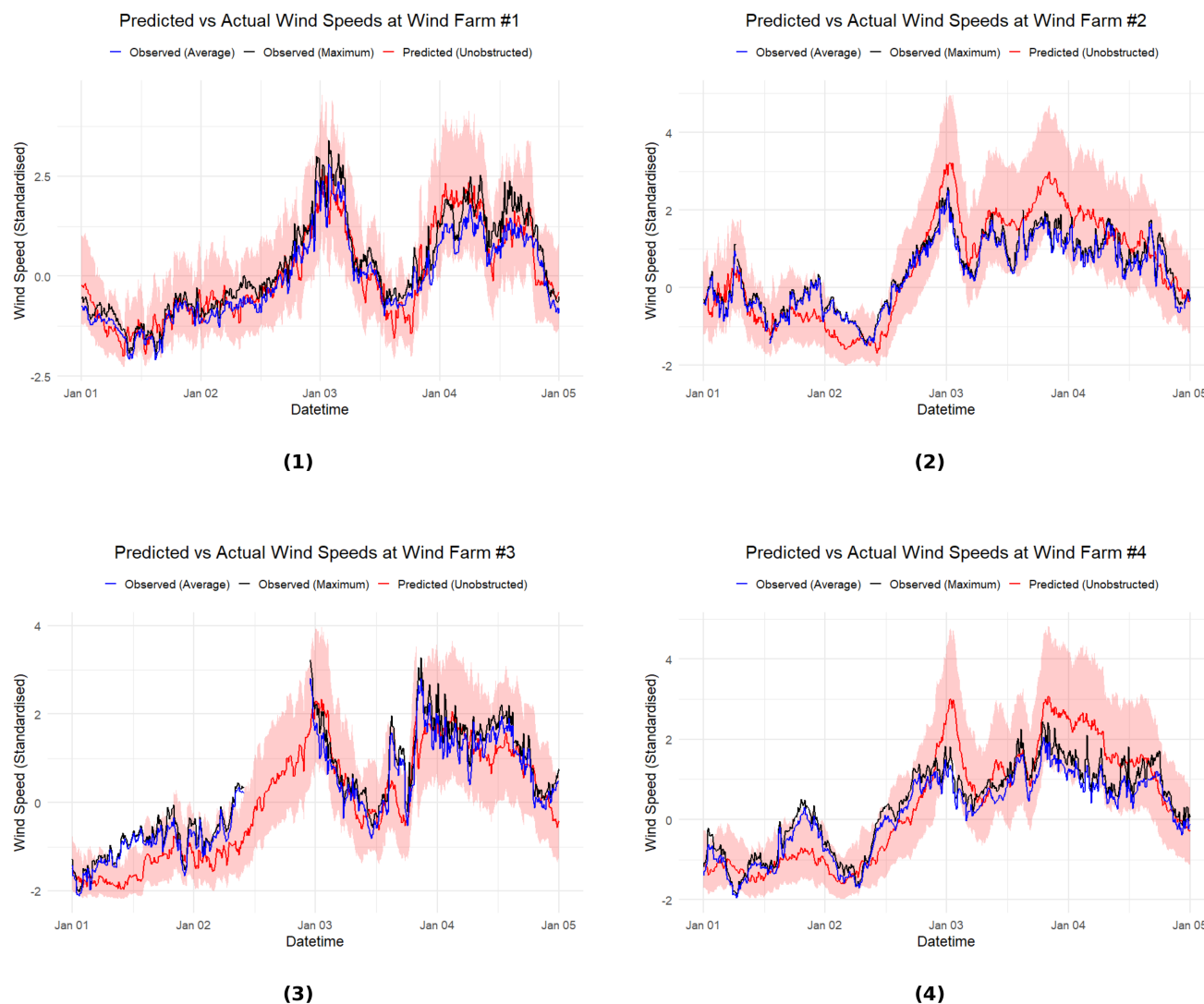


Figure 13. Predicted and observed wind speeds at four wind farms over the first four days of the year (values have been standardised for data privacy). Shaded regions denote the 95% prediction interval. In each panel, the average wind speed is shown in blue and the maximum wind speed in black. Periods with missing turbine observations are also visible.



Farm	RMSE (m/s)		Bias (m/s)		Corr.	
	Proposed	ERA5	Proposed	ERA5	Proposed	ERA5
WF1	1.97	2.20	0.31	1.27	0.87	0.89
WF2	1.89	1.4	-0.85	0.30	0.88	0.90
WF3	1.84	3.01	0.20	2.57	0.86	0.90
WF4	1.98	1.86	-0.52	-0.18	0.85	0.86
WF5	1.73	1.63	-0.1	1.09	0.88	0.92
WF6	1.78	2.44	-0.25	1.92	0.88	0.92
WF7	2.08	2.00	-0.97	1.33	0.84	0.87

Table 4. Prediction accuracy at each individual wind farm, evaluated against the **maximum** observed wind speed (proxy for unobstructed wind speed). Mean bias is defined as observed minus predicted, so positive values indicate underprediction and negative values indicate overprediction.

Model	RMSE (m/s)	Mean Bias (m/s)	Correlation
Proposed Model (Unadjusted)	2.01	-1.13	0.86
ERA5	1.72	0.38	0.86
Proposed Model (10% Wake Loss)	1.70	-0.29	0.86
Proposed Model (15% Wake Loss)	1.62	0.14	0.86
Proposed Model (20% Wake Loss)	1.67	0.56	0.86

Table 5. Prediction accuracy aggregated across all seven wind farms, evaluated against the **average** observed wind speed at each farm.

440 Finally, we validate our uncertainty quantification by examining the empirical coverage of our prediction intervals (Table 6), which calculates what percentage of true values are contained within a given confidence interval. The 80% and 95% intervals show coverage of 85% and 96.6%, respectively, indicating that the estimated variance is slightly too high but well-calibrated.

Prediction Interval	Empirical Coverage
80%	85.0%
95%	96.6%

Table 6. Empirical coverage of the prediction intervals at wind farms of our proposed model, demonstrating well-calibrated uncertainty quantification. These values are calculated using the maximum observed wind speed at all 7 validation sites, at each 10-minute time point over the full year.



5 Conclusions

This paper presented a two-step statistical framework to predict wind speeds at turbine hub heights using publicly available meteorological data from separate locations. By combining a non-parametric height extrapolation model with a spatial Gaussian Process, our approach produces high-resolution wind speed estimates at wind farm locations. When validated against seven onshore wind farms in Ireland, our model achieves comparable accuracy to ERA5 while being available in near real-time, demonstrating its suitability for operational use.

Given the open-access nature of the input data and prediction framework, this could be used as a tool alongside reanalysis data for preliminary site selection. Open-access wind resource tools can be important for smaller scale community energy groups who don't have the up front investment for extensive site surveys (Araveti et al., 2022). The framework provides a valuable tool for short-term wind monitoring and real-time decision-making. Unlike ERA5, which is available with a 5-day delay, our approach can deliver immediate estimates, enabling grid operators to adjust reserve capacity or optimize dispatch schedules (Wang et al., 2016). In addition, the framework generates continuous 10-minute resolution time series at any prospective wind farm location, enabling both real-time monitoring and retrospective analysis at spatial scales not directly available from reanalysis products. The provision of uncertainty estimates at each site further allows stakeholders to quantify the confidence in predictions and integrate probabilistic forecasts into operational planning, which is crucial for managing variability in wind generation.

Our model is sensitive to the choice of height extrapolation and the underlying reanalysis dataset. Non-parametric methods such as the Generalized Additive Model (GAM) used here allow flexibility in capturing local wind profiles, but performance may vary in regions with complex terrain, high forest cover, or limited data coverage. The use of reanalysis datasets inherently smooths local variability, which can under-represent extreme or highly localized wind events. Future work could explore the fusion of multiple reanalysis products or satellite-derived wind estimates to enhance model robustness (Gruber et al., 2022; Dörenkämper et al., 2020).

While we focus on a two step approach, a joint multivariate model of wind speeds across multiple heights is conceivable. This strategy is impractical for our use case because (i) no hub-height data are available for conditioning the model during application, and (ii) hub heights vary between farms, complicating a unified approach. However for practitioners who have access to both meteorological station and hub-height observations, a multivariate approach, such as a coregionalisation model (Krainski et al., 2018), or multi-output GP could be employed (Alvarez and Lawrence, 2008).

Another limitation lies in the treatment of wake effects within wind farms. While we apply empirical reduction factors to account for speed losses at downwind turbines, more detailed modelling could incorporate turbine layouts and computational fluid dynamics (CFD) simulations (González-Longatt et al., 2012; Pryor and Barthelmie, 2024). Such refinements would be particularly relevant when the model is applied to forecast farm-level energy output or turbine loads.

Beyond onshore applications, the proposed framework is naturally extendable to offshore wind estimation, where observational networks are typically much sparser and direct hub-height measurements are rarely available. In this setting, the 10 m reference observations used in the first stage of the framework could be supplied by offshore weather buoys, coastal



meteorological stations, satellite-derived surface winds, or numerical weather prediction products. For example, Advanced Scatterometer (ASCAT) products provide near-real-time offshore wind estimates at approximately 10 m above sea level and have previously been evaluated in Irish waters (Remmers et al., 2019). Since such products do not directly provide turbine
 480 hub-height winds, they could be used as inputs to the height-extrapolation stage of the proposed framework, with the resulting hub-height estimates then spatially interpolated to offshore wind farm locations.

However, a rigorous offshore application would require additional validation data, such as offshore meteorological masts, floating LiDAR measurements, or operational offshore wind farm observations at hub height. These data were not available for the present study, which is therefore restricted to validation at onshore wind farms. The offshore extension should consequently
 485 be viewed as a promising application of the modelling framework rather than a demonstrated result. In particular, offshore conditions may require re-training or recalibration of the height-extrapolation model, since vertical wind profiles over sea differ from those over land due to differences in surface roughness, atmospheric stability, and boundary-layer structure. Future work could therefore combine buoy, ASCAT, and numerical weather prediction inputs with offshore hub-height validation data to assess whether the same two-stage strategy can provide real-time, sub-hourly wind-speed estimates for offshore wind
 490 resource assessment and short-term operational monitoring.

Combining our statistical framework with numerical weather prediction (NWP) outputs may further improve sub-hourly forecasting, addressing a key limitation of traditional NWP models, which typically provide data at 10 m heights (Olauson, 2018; Gualtieri, 2019).

From a practical perspective, this framework could reduce reliance on expensive on-site wind measurements for preliminary
 495 assessments. Traditional wind resource assessments require extended campaigns, often costing tens of thousands of euros per site (Teagasc, 2020), whereas our approach leverages open-access datasets to provide rapid, high-resolution estimates. This has clear implications for project development, investment decisions, and grid management, particularly when considering multiple candidate sites or monitoring operating wind farms in real time.

In summary, our two-step approach enables high-resolution, real-time predictions of turbine-height wind speeds from pub-
 500 licly available data, with quantified uncertainty. While the method performs comparably to ERA5 in an Irish context, future work should focus on improving wake effect modelling, exploring multi-source data fusion, and extending the framework to offshore applications to maximize its practical impact.

Code and data availability. Reanalysis and meteorological station data is publicly available. Data from the Global Wind Atlas can be down-
 loaded at <https://globalwindatlas.info>. Data from the New European Wind Atlas can be downloaded at www.neweuropeanwindatlas.eu. ERA5
 505 data is available at cds.climate.copernicus.eu. Meteorological station is maintained by Met Éireann, the Irish national meteorological service. Historical data is available at www.met.ie/climate/available-data/historical-data. The validation dataset was shared by an industry partner and is not publicly available. Code for final models and data processing will be available at: <https://github.com/EamonnO22>.



510 *Author contributions.* All authors contributed to concept development, writing and editing. EO led core research tasks including model development, software development, testing and benchmarking. MU, DA, LB and JS contributed to problem formulation, metrics, and statistical design.

Competing interests. The authors do not have competing interests to report.

515 *Acknowledgements.* This publication has emanated from research supported in part by a grant from Taighde Éireann – Research Ireland under Grant number 18/CRT/6049. This publication has emanated from research conducted with the financial support of the EU Commission Recovery and Resilience Facility under the Research Ireland Energy Innovation Challenge Grant Number 22/NCF/EI/11162G. This work was partially supported by the “PHC ULYSSES” program (project number: 50252NC), funded by the French Ministry for Europe and Foreign Affairs, the French Ministry for Higher Education and Research, Taighde Eireann – Research Ireland. The French authors also acknowledge the financial support of the Chair Geolearning, funded by ANDRA, BNP Paribas, CCR and the SCOR Foundation for Science.

520 Data obtained from the Global Wind Atlas version 4.0, a free, web-based application developed, owned and operated by the Technical University of Denmark (DTU). The Global Wind Atlas version 4.0 is released in partnership with the World Bank Group, utilizing data provided by Vortex, using funding provided by the Energy Sector Management Assistance Program (ESMAP). For additional information: <https://globalwindatlas.info>

Data is also obtained from the New European Wind Atlas, a free, web-based application developed, owned and operated by the NEWA Consortium. For additional information see www.neweuropeanwindatlas.eu

The authors thank Alejandro Fernandez and Greencoat Renewables for providing the wind farm data used for validation in this study.



525 References

- Adaramola, M. and Krogstad, P.-Å.: Experimental investigation of wake effects on wind turbine performance, *Renewable energy*, 36, 2078–2086, <https://doi.org/https://doi.org/10.1016/j.renene.2011.01.024>, 2011.
- Alvarez, M. and Lawrence, N.: Sparse convolved Gaussian processes for multi-output regression, *Advances in neural information processing systems*, 21, https://proceedings.neurips.cc/paper_files/paper/2008/file/149e9677a5989fd342ae44213df68868-Paper.pdf, 2008.
- 530 Araveti, S., Quintana, C. A., Kairisa, E., Mutule, A., Adiazola, J. P. S., Sweeney, C., and Carroll, P.: Wind energy assessment for renewable energy communities, *Wind*, 2, 325–347, <https://doi.org/https://doi.org/10.3390/wind2020018>, 2022.
- Bailey, B. H., McDonald, S. L., Bernadett, D. W., Markus, M. J., and Elsholz, K. V.: Wind resource assessment handbook: Fundamentals for conducting a successful monitoring program, Tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States); AWS . . . , <https://doi.org/https://doi.org/10.2172/486127>, 1997.
- 535 Banerjee, S., Carlin, B. P., and Gelfand, A. E.: Hierarchical modeling and analysis for spatial data, Chapman and Hall/CRC, 2003.
- Bowden, G., Barker, P., Shestopal, V., and Twidell, J.: The Weibull distribution function and wind power statistics, *Wind Engineering*, pp. 85–98, <https://ui.adsabs.harvard.edu/abs/1983WiEng...7...85B>, 1983.
- Carta, J. A., Ramirez, P., and Velazquez, S.: A review of wind speed probability distributions used in wind energy analysis: Case studies in the Canary Islands, *Renewable and sustainable energy reviews*, 13, 933–955, <https://doi.org/10.1016/j.rser.2008.05.005>, 2009.
- 540 Crippa, P., Alifa, M., Bolster, D., Genton, M. G., and Castruccio, S.: A temporal model for vertical extrapolation of wind speed and wind energy assessment, *Applied Energy*, 301, 117 378, <https://doi.org/https://doi.org/10.1016/j.apenergy.2021.117378>, 2021.
- Cui, L. and Xie, M.: Some normal approximations for renewal function of large Weibull shape parameter, *Communications in Statistics-Simulation and Computation*, 32, 1–16, <https://doi.org/https://doi.org/10.1081/SAC-120013107>, 2003.
- Dai, A. and Deser, C.: Diurnal and semidiurnal variations in global surface wind and divergence fields, *Journal of Geophysical Research: Atmospheres*, 104, 31 109–31 125, <https://doi.org/10.1029/1999jd900927>, 1999.
- 545 Davis, N. N., Badger, J., Hahmann, A. N., Hansen, B. O., Mortensen, N. G., Kelly, M., Larsén, X. G., Olsen, B. T., Floors, R., Lizcano, G., et al.: The Global Wind Atlas: A high-resolution dataset of climatologies and associated web-based application, *Bulletin of the American Meteorological Society*, 104, E1507–E1525, <https://doi.org/10.1175/BAMS-D-21-0075.1>, 2023.
- Department of the Environment, Climate and Communications: Irish Government Climate Action Plan, <https://www.gov.ie/en/publication/79659-climate-action-plan-2024/>, accessed: 2024-11-18, 2024.
- 550 Doddy Clarke, E., Griffin, S., McDermott, F., Monteiro Correia, J., and Sweeney, C.: Which reanalysis dataset should we use for renewable energy analysis in Ireland?, *Atmosphere*, 12, 624, <https://doi.org/10.3390/atmos12050624>, 2021.
- Doohan, J., Kook, L., and Burke, K.: Comparison of generalised additive models and neural networks in applications: A systematic review, *Expert Systems with Applications*, p. 131082, <https://doi.org/10.1016/j.eswa.2025.131082>, 2026.
- 555 Dörenkämper, M., Olsen, B. T., Witha, B., Hahmann, A. N., Davis, N. N., Barcons, J., Ezber, Y., García-Bustamante, E., González-Rouco, J. F., Navarro, J., et al.: The making of the new european wind atlas—part 2: Production and evaluation, *Geoscientific Model Development Discussions*, 2020, 1–37, <https://doi.org/https://doi.org/10.5194/gmd-13-5079-2020>, 2020.
- Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., et al.: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), *Journal of climate*, 30, 5419–5454, <https://doi.org/10.1175/JCLI-D-16-0758.1>, 2017.
- 560 Global Wind Atlas: GWA Release Notes, <https://globalwindatlas.info/en/about/ReleaseNotes>, accessed: 2025-12-01, 2025.



- González-Longatt, F., Wall, P., and Terzija, V.: Wake effect in wind farm performance: Steady-state and dynamic behavior, *Renewable Energy*, 39, 329–338, <https://doi.org/https://doi.org/10.1016/j.renene.2011.08.053>, 2012.
- Gruber, K., Klöckl, C., Regner, P., Baumgartner, J., and Schmidt, J.: Assessing the Global Wind Atlas and local measurements for bias correction of wind power generation simulated from MERRA-2 in Brazil, *Energy*, 189, 116212, <https://doi.org/https://doi.org/10.1016/j.energy.2019.116212>, 2019.
- Gruber, K., Regner, P., Wehrle, S., Zeyringer, M., and Schmidt, J.: Towards global validation of wind power simulations: A multi-country assessment of wind power simulation from MERRA-2 and ERA-5 reanalyses bias-corrected with the global wind atlas, *Energy*, 238, 121520, <https://doi.org/https://doi.org/10.1016/j.energy.2021.121520>, 2022.
- Gualtieri, G.: A comprehensive review on wind resource extrapolation models applied in wind energy, *Renewable and Sustainable Energy Reviews*, 102, 215–233, <https://doi.org/https://doi.org/10.1016/j.rser.2018.12.015>, 2019.
- Gualtieri, G.: Analysing the uncertainties of reanalysis data used for wind resource assessment: A critical review, *Renewable and Sustainable Energy Reviews*, 167, 112741, <https://doi.org/https://doi.org/10.1016/j.rser.2022.112741>, 2022.
- Hahmann, A. N., Sile, T., Witha, B., Davis, N. N., Dörenkämper, M., Ezber, Y., García-Bustamante, E., González Rouco, J. F., Navarro, J., Olsen, B. T., et al.: The making of the new European Wind Atlas, Part 1: model sensitivity, *Geoscientific Model Development Discussions*, 2020, 1–33, <https://doi.org/https://doi.org/10.5194/gmd-13-5053-2020>, 2020.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, *Quarterly journal of the royal meteorological society*, 146, 1999–2049, <https://doi.org/10.1002/qj.3803>, 2020.
- Houndekind, F. and Ouarda, T. B.: Machine learning and statistical approaches for wind speed estimation at partially sampled and unsampled locations; review and open questions, *Energy Conversion and Management*, 327, 119555, <https://doi.org/https://doi.org/10.1016/j.enconman.2025.119555>, 2025.
- Krainski, E., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., and Rue, H.: Chapter 3: More than one likelihood, in: *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA*, chap. 3, Chapman and Hall/CRC, <https://doi.org/10.1201/9780429031892>, 2018.
- Lahoz, W. A. and Schneider, P.: Data assimilation: making sense of Earth Observation, *Frontiers in Environmental Science*, 2, 16, <https://doi.org/10.3389/fenvs.2014.00016>, 2014.
- Lenzi, A. and Genton, M. G.: Spatiotemporal probabilistic wind vector forecasting over Saudi Arabia, *The Annals of Applied Statistics*, 14, 1359–1378, <https://doi.org/10.1214/20-AOAS1347>, 2020.
- Mc Garrigle, E. V. and Leahy, P. G.: Quantifying the value of improved wind energy forecasts in a pool-based electricity market, *Renewable Energy*, 80, 517–524, <https://doi.org/https://doi.org/10.1016/j.renene.2015.02.023>, 2015.
- Met Éireann: Met Éireann’s Wind Measurement, <https://www.met.ie/climate/what-we-measure/wind>, accessed: 2025-03-18, 2025.
- Michelangeli, P.-A., Vrac, M., and Loukos, H.: Probabilistic downscaling approaches: Application to wind cumulative distribution functions, *Geophysical Research Letters*, 36, <https://doi.org/https://doi.org/10.1029/2009GL038401>, 2009.
- Murthy, K. and Rahi, O.: A comprehensive review of wind resource assessment, *Renewable and Sustainable Energy Reviews*, 72, 1320–1342, <https://doi.org/10.1016/j.rser.2016.10.038>, 2017.
- Oke, T.: Guide to instruments and methods of observation, WMO: Geneva, Switzerland, 3, 426, <https://library.wmo.int/idurl/4/41650>, 2018.
- Olauson, J.: ERA5: The new champion of wind power modelling?, *Renewable energy*, 126, 322–331, <https://doi.org/https://doi.org/10.1016/j.renene.2018.03.056>, 2018.



- Organ, E., Upton, M., Allard, D., Benoit, L., and Sweeney, J.: Enhancing the Accuracy of Spatio-Temporal Models for Wind Speed Prediction
 600 by Incorporating Bias-Corrected Crowdsourced Data, *Environmetrics*, 37, e70 069, <https://doi.org/10.1002/env.70069>, 2026.
- Panofsky, H. A. and Dutton, J. A.: Atmospheric turbulence. Models and methods for engineering applications, New York: Wiley, <https://ui.adsabs.harvard.edu/abs/1984atmm.book.....P/abstract>, 1984.
- Pryor, S. C. and Barthelmie, R. J.: Wind shadows impact planning of large offshore wind farms, *Applied Energy*, 359, 122 755, <https://doi.org/10.1016/j.apenergy.2024.122755>, 2024.
- 605 Ramon, J., Lledó, L., Torralba, V., Soret, A., and Doblas-Reyes, F. J.: What global reanalysis best represents near-surface winds?, *Quarterly Journal of the Royal Meteorological Society*, 145, 3236–3251, <https://doi.org/10.1002/qj.3616>, 2019.
- Remmers, T., Cawkwell, F., Desmond, C., Murphy, J., and Politi, E.: The potential of advanced scatterometer (ASCAT) 12.5 km coastal observations for offshore wind farm site selection in Irish waters, *Energies*, 12, 206, <https://doi.org/10.3390/en12020206>, 2019.
- 610 Rue, H. and Held, L.: Gaussian Markov random fields: theory and applications, Chapman and Hall/CRC, <https://doi.org/10.1201/9780203492024>, 2005.
- Schlez, W. and Infield, D.: Horizontal, two point coherence for separations greater than the measurement height, *Boundary-Layer Meteorology*, 87, 459–480, <https://doi.org/10.1023/A:1000997610233>, 1998.
- Sustainable Energy Authority of Ireland: SEAI Wind Atlas, <https://maps.seai.ie/apps/WindAtlas/>, accessed: 2025-11-06, 2025a.
- 615 Sustainable Energy Authority of Ireland: Wind Farms in Ireland, <https://data.gov.ie/dataset/wind-farms-in-ireland>, accessed: 2025-10-20, 2025b.
- Sweeney, C., Bessa, R. J., Browell, J., and Pinson, P.: The future of forecasting for renewable energy, *Wiley Interdisciplinary Reviews: Energy and Environment*, 9, e365, <https://doi.org/10.1002/wene.365>, 2020.
- Teagasc: Wind Energy Fact Sheet, <https://teagasc.ie/wp-content/uploads/2025/05/Energy-11-Wind-Energy.pdf>, accessed: 2025-10-29, 2020.
- 620 Vassallo, D., Krishnamurthy, R., and Fernando, H. J.: Decreasing wind speed extrapolation error via domain-specific feature extraction and selection, *Wind Energy Science*, 5, 959–975, <https://doi.org/10.5194/wes-5-959-2020>, 2020.
- Verhoef, A., Portabella, M., and Stoffelen, A.: High-resolution ASCAT scatterometer winds near the coast, *IEEE Transactions on Geoscience and Remote Sensing*, 50, 2481–2487, <https://doi.org/10.1109/TGRS.2011.2175001>, 2012.
- Vest, K. and Tych, W.: Is the apparent global stilling effect on wind power generation an artefact of sampling rate? Evidence from high-
 625 frequency observations in the UK, *Renewable Energy*, p. 125097, <https://doi.org/10.1016/j.renene.2025.125097>, 2025.
- Wais, P.: A review of Weibull functions in wind sector, *Renewable and Sustainable Energy Reviews*, 70, 1099–1107, <https://doi.org/10.1016/j.rser.2016.12.014>, 2017.
- Wang, Q., Wu, H., Florita, A. R., Martinez-Anido, C. B., and Hodge, B.-M.: The value of improved wind power forecasting: Grid flexibility quantification, ramp capability analysis, and impacts of electricity market operation timescales, *Applied Energy*, 184, 696–713, <https://doi.org/10.1016/j.apenergy.2016.11.016>, 2016.
- 630 Weekes, S. M. and Tomlin, A. S.: Low-cost wind resource assessment for small-scale turbine installations using site pre-screening and short-term wind measurements, *IET Renewable Power Generation*, 8, 349–358, <https://doi.org/10.1049/iet-rpg.2013.0152>, 2014.
- Williams, C. K. and Rasmussen, C. E.: Gaussian processes for machine learning, MIT press Cambridge, MA, <https://doi.org/10.7551/mitpress/3206.001.0001>, 2006.



Wind Energy Ireland: 2025 Annual report, <https://blog.windenergyireland.com/irish-wind-farms-provided-a-third-of-our-power-in-2025>, accessed: 2026-07-01, 2026.

Wood, S. N.: Generalized additive models: an introduction with R, chapman and hall/CRC, <https://doi.org/https://doi.org/10.1201/9781315370279>, 2017.

640 Yu, S. and Vautard, R.: A transfer method to estimate hub-height wind speed from 10 meters wind speed based on machine learning, Renewable and Sustainable Energy Reviews, 169, 112 897, <https://doi.org/https://doi.org/10.1016/j.rser.2022.112897>, 2022.