Correlating wake-model performance to the atmospheric state: a SCADA-based analysis at two offshore wind farms

Maxime Lejeune[1], Lukas Vollmer[2], Nicolas Coudou[1], Leonardo Alcayaga[3], and Wim Munters[1]

[1]Environmental and Applied Fluid Dynamics, von Karman Institute for Fluid Dynamics, Sint-Genesius-Rode, Belgium
[2]Fraunhofer Institute for Wind Energy Systems (IWES), Oldenburg, Germany
[3]DTU Wind and Energy Systems, Technical University of Denmark, Risø, Denmark

Correspondence: Maxime Lejeune (maxime.lejeune@vki.ac.be)

Abstract. Engineering wake models are routinely validated against SCADA measurements with discrepancies between predicted and observed power often attributed directly to limitations in the wake model itself or to inadequate calibration. Wind farm power assessment however involves a sequence of processing steps, of which wake modelling is only the final stage. Disagreement between predictions and observations may therefore stem from the wake-model physics itself but also from the software implementation, the reconstructed inflow, or the atmospheric conditions sampled.

This work aims to assess these different contributions independently by evaluating the entire wake modelling chain, from wind-resource assessment to the final power estimation. SCADA data from two offshore wind farms, Westermost Rough and Anholt, are processed using two independently developed filtering workflows and complemented with profiling LiDAR measurements and 1 km-resolution WRF simulations. The resulting inflow conditions are used to drive two wake models, Jensen and TurbOPark, both sensitized to the ambient turbulence intensity. Three wake-modelling software packages are considered (FLORIS, PyWake, and FOXES) while WindIO is used to ensure identical inputs across all implementations.

The two SCADA-filtering workflows yield nearly interchangeable wind speed timeseries (bias $\leq 0.04\,\mathrm{m\,s^{-1}}$, $r \approx 1$), but differ in wind direction by up to 7.27° at Anholt, likely due to differences in northing calibration procedure. For identical inputs, software implementation alone produces bias spreads of up to $4\,\%$ for TurbOPark, comparable to the effect of the wake-model choice itself, while Jensen remains consistent across packages. Farm-level bias is further shown to be an insufficient performance indicator that should therefore be complemented by additional metrics such as farm-level RMSE, correlation, and turbine-level statistics. Indeed, bias compensation can lead to near-zero farm-wide biases that hide binned farm-mean errors of $\pm 5\,\%$ and farm-MAE exceeding $20\,\%$ under very stable or strongly heterogeneous inflow conditions. At the turbine-level, the analysis highlights the need to account for local wind speed gradients when evaluating individual turbine power, establishing clear spatial correlation between wind speedup maps and turbine power error spatial distribution.

Overall, these results support a shift away from AEP-based validation toward time-series or condition-binned approaches. They additionally motivate a better inclusion of physical phenomena such as heterogeneity-aware models. They finally call for better community standards, through, *e.g.* the development of shared standards such as WindIO.

Copyright statement.

1 Introduction

Wind turbine wakes are the primary source of power losses in wind farms, with Annual Energy Production (AEP) losses as high as 20 % for offshore sites (Lee and Fields, 2021). These wakes not only affect turbines operating within the same farm but can propagate over extended distances, leading to significant inter-farm losses. Accurate prediction of these losses is therefore a prerequisite for reliable wind farm design, yield assessment, and operational planning, and carries direct implications for the economics of offshore wind (O'Malley et al., 2024).

Despite the recent gain of popularity of large-eddy simulations (Jong et al., 2026) and mesoscale models (Rosencrans et al., 2024; Porchetta et al., 2026), steady-state engineering wake models often remain the tool of choice for wake loss assessment. These models typically describe the wake wind speed deficit as an analytical expression fitted to site data. Their conceptual simplicity and low computational cost make them uniquely suited for applications requiring a high number of evaluations, such as iterative energy yield assessment. Over the past decades, a rich landscape of wake deficit models has been developed, including the classical top-hat deficit formulation of Jensen (1983), the Gaussian formulation derived by Bastankhah and Porté-Agel (2014), or the TurbOPark model (Pedersen et al., 2022). In parallel, efforts have focused on better capturing deep-array effects through improved wake superposition schemes (Zong and Porté-Agel, 2020; Lanzilao and Meyers, 2021) and parametrisations for wake-added turbulence (Crespo and Hernández, 1996). These models have progressively been assembled into both open-source toolboxes - such as FLORIS(NREL, 2025), PyWake(Pedersen et al., 2023) and FOXES(Schmidt et al., 2023) - and commercial packages, contributing to their widespread adoption.

Over the years, wake models have been validated through numerous studies taking advantage of the growing availability of Supervisory Control And Data Acquisition (SCADA) data, offering unprecedented insight into wind farm physics. An early comparison led by Barthelmie and Jensen (2010) at the Nysted offshore wind farm revealed that engineering wake models systematically underestimate deep-array wake losses, though Gaumond et al. (2014) showed that properly accounting for wind direction uncertainty closes much of the apparent model-observation gap. Hansen et al. (2012) further demonstrated a clear dependence of wake-induced power deficits on atmospheric stability. More recently, Nygaard et al. (2022) validated TurbOPark and Park against SCADA data from 19 offshore wind farms, quantifying their respective uncertainties and biases. Vollmer et al. (2024) extended the benchmarking effort to farm-to-farm interactions at Sandbank and Dan Tysk, showing that engineering models approximate overall wake losses well but systematically underestimate the increased losses under stable conditions. This validation effort has gone hand in hand with a quest for optimal recalibrated tuning parameters, ultimately leading to the proposition of condition-dependent (Van Binsbergen et al., 2024) or even dynamic (Doekemeijer et al., 2019) wake model calibration frameworks.

Indeed, the parameters of engineering wake models, often treated as universal, are in fact sensitive to the meteorological conditions at the site. Atmospheric stability and Turbulence Intensity (TI) govern wake recovery rates (Hansen et al., 2012; Barthelmie and Jensen, 2010), with stable stratification producing longer, more persistent wakes (Platis et al., 2018, 2022), and

seasonal or synoptic variability adding further structure (Rosencrans et al., 2024; Palatos-Plexidas et al., 2024). Additionally, at coastal sites such as Anholt or Westermost Rough, inflow heterogeneity driven by land-sea gradients introduces another layer of spatial complexity as reported by, *e.g.* , van der Laan et al. (2017), Ahsbahs et al. (2018) or Peña et al. (2018) for the Anholt wind farm. Nygaard et al. (2022) additionally state that accounting for inflow heterogeneity is an essential component for a fair model comparison based on 19 offshore wind farms benchmark. These sensitivities imply that model performances are inherently site- and condition-dependent, and that a single set of parameters cannot be expected to yield consistent accuracy across diverse sites and atmospheric regimes.

A second, less recognised source of variability arises not from physics but from software. Mudafort et al. (2024) showed that even when using the same wake model formulation, implementation choices, such as rotor-averaging methods, wake combination schemes, and near-wake treatment, introduce differences across software packages. This raises a fundamental question for validation studies: when a model disagrees with observations, how much of the discrepancy originates from model physics, how much from user's implementation choices, and how much from the inflow and observational data used as inputs?

Addressing these questions requires to consider the entire wake modelling workflow and the subsequent validation procedure, not just the wake model itself. Wake modelling for wind farm power estimation typically involves two main stages: wind resource assessment (WRA) and the wake modelling itself. WRA aims at characterising the inflow conditions. Its complexity greatly varies depending on the data source: extracting point data from mesoscale simulations is relatively straightforward, though it may require additional bias correction (Hahmann et al., 2020), while estimating the inflow state from SCADA measurements requires extensive filtering to remove faulty measurements, curtailment periods, or abnormal operating conditions. Light Detection And Ranging (LiDAR) systems complete the picture by resolving the flow over the full rotor-swept height, and are now mature enough to serve as reference instruments (Clifton et al., 2018). Yet, they remain spatially punctual, weather-dependent, and subjected to well-documented biases in the retrieved turbulence statistic (Kelberlau and Mann, 2020). Regardless of the data source, the modeller must make a series of informed choices when designing the setup, for example, which thresholds to use for flagging abnormal turbine operation or how to filter the data temporally.

The accuracy of the resulting power estimates depends not only on the WRA methodology but also on the engineering wake model itself, which is subject to its own limitations as discussed above. Each step of the workflow therefore introduces additional sources of uncertainty depending on the modeller's choices and the limitations of the tools used. In this work, we distinguish four main source of uncertainties:

- *Inflow uncertainties*: variations in derived inflow conditions resulting from different WRA strategies, in particular from different SCADA-processing workflows;
- *Wake-model uncertainties*: differences in power predictions between wake models;
- *Implementation uncertainties*: differences arising from distinct software implementations of the same wake model;
- *Condition-dependent uncertainties*: differences in wake-model accuracy depending on the atmospheric state, such as the evolution of power prediction bias across stability classes.

Most existing validation studies address these workflow layers simultaneously so that their relative contribution to the global error cannot be evaluated separately. A first step toward a more systematic decomposition was made by Lejeune et al. (2026), who applied two independent SCADA filtering workflows and three software implementations of two wake models to the Westermost Rough offshore wind farm. That study found TurbOPark AEP spreads of up to $\pm 1\,\%$ across software for identical inputs, and direction-dependent power discrepancies of 1-2 % between filtering workflows. It further showed that the Weather Research and Forecasting (WRF) mesoscale model provides a reasonable mean inflow estimate but lacks the temporal accuracy of SCADA-derived conditions, and that the wake model error correlates with atmospheric stability and wind direction. The analysis was, however, limited to a single wind farm, two wake models operating at fixed ambient TI, and a single year of data.

The present paper extends this work along three dimensions. First, the analysis is expanded both spatially and temporally: the Anholt wind farm is included in the study while the Westermost Rough analysis is revisited, extending it to two years of data. This paper consequently covers contrasting North Sea and Kattegat environments with different turbine types and layouts. Second, the wake models are sensitized to the ambient TI, allowing them to better respond to changes in atmospheric state compared to the original study where the TI was initially held constant. Third, the contextual analysis of prediction errors is extended to include horizontal flow heterogeneity, highlighting its intricate interplay with wake recovery. Together, this extended analysis shall allow to gain a better understanding of the uncertainties involved in wake model validation with SCADA data by systematically decomposing the different steps of the workflow.

The remainder of the paper is structured as follows. Section 2 describes the SCADA, LiDAR, and WRF datasets for both wind farms. Section 3 details the SCADA filtering workflows and the wake-modelling framework. Section 4 analyses and compares the inflow conditions derived from the different sources. Section 5 presents the wake modelling results, including the impact of atmospheric stability and flow heterogeneity on prediction errors. Section 6 discusses the impact of the different sources of uncertainties on the overall wake modeling workflow accuracy, and Section 7 summarises the main conclusions.

2 Dataset descriptions

This paper studies two offshore wind farms operated by Ørsted: Westermost Rough and Anholt. Figure 1 shows the layout of both farms along with the mean hub-height wind speed field retrieved from WRF (see Sect. 2.3 below for a comprehensive description of the WRF setup).

Westermost Rough is located off the east coast of the UK and consists of 35 Siemens Swt-6.0-154 turbines. With a hub height of 102 m, each turbine has a nameplate capacity of 6.0 MW, summing up to a total installed capacity of approximately 210 MW. The coastline lies approximately 7 km to the west, inducing strong coastal velocity gradients for westerly winds as visible in Fig. 1. To the east, a large open-sea fetch extends unobstructed. Westermost Rough was additionally built in the vicinity of the Humber Gateway wind farm, located to the south-east and consisting of 73 Vestas V112-3.0 MW turbines totaling 219 MW. In addition to the LiDAR and SCADA data provided by Ørsted, this study makes use of 1 km-resolution WRF mesoscale simulations conducted within the scope of the EU Horizon project DTWO (DTWO Consortium, 2024; Larsén and the DTWO Consortium, 2025).

Anholt is located in the Kattegat, between the Djursland peninsula and Anholt island, approximately 20 km from shore. With 111 Siemens SWT-3.6-120 turbines and a total installed capacity of 400 MW, it was the largest wind farm in Denmark until 2019. The Siemens SWT-3.6-120 has a rated power of 3.6 MW and a hub height of 82 m. To the west of the wind farm lies the Danish coast, while an approximately 100 km long open-sea fetch extends to the north-east with the Anholt island the only obstacle. As for Westermost Rough, both site measurements (SCADA and LiDAR) and WRF mesoscale data are used.

The remainder of this section describes the different datasets used: field data (SCADA and LiDAR) is first introduced, after which the WRF simulation framework is presented.

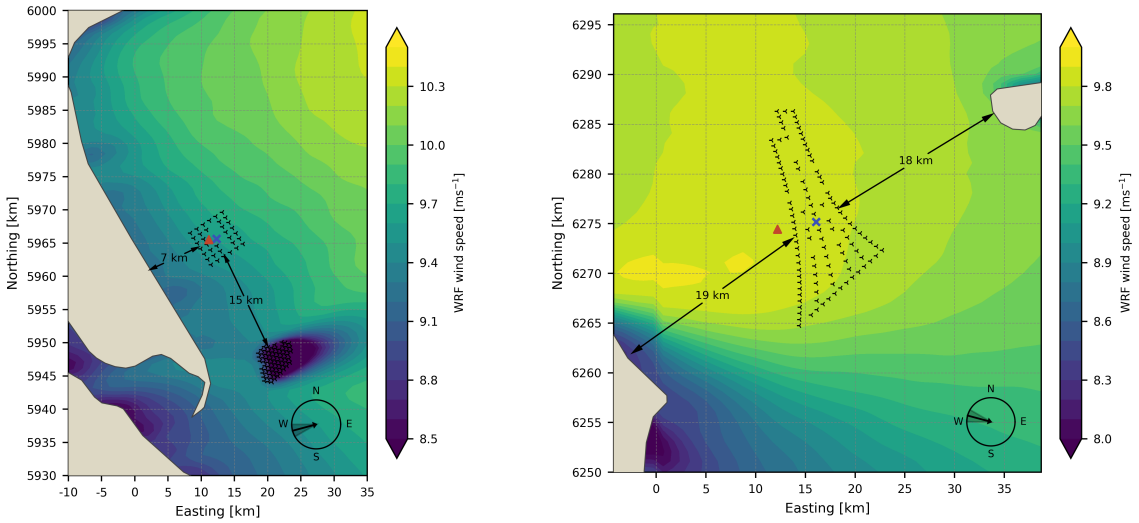


Figure 1. Wind farm layouts: mean hub-height WRF wind speed field for the dominant wind direction $\pm 15^\circ$ (255° for Westermost Rough and 285° for Anholt) with turbine positions overlaid. **(Left)** Westermost Rough, and the neighboring Humber Gateway wind farm parametrized using Fitch et al. (2012). **(Right)** Anholt. [▲] LiDAR position. [×] Farm barycenter.

2.1 SCADA datasets

The SCADA data were provided by Ørsted in the form of 10-minute aggregated records for both wind farms. This raw SCADA data was later processed through two independent SCADA-filtering workflows as discussed in Section 3.1. These workflows are named after the institutions that developed them, namely the *VKI* and *IWES workflows*.

For Westermost Rough, the dataset covers the complete calendar years 2016 and 2017. For Anholt, the dataset spans September 2013 to June 2015, with varying data availability across the period.

For each turbine, the following channels are used in the present study: (i) *mean active power output*, $\bar{P}_{10\,\mathrm{min}}$ (ii) *mean nacelle anemometer wind speed*, $\bar{u}_{10\,\mathrm{min}}$, (iii) *standard deviation of nacelle anemometer wind speed*, $\sigma_{u,10\,\mathrm{min}}$, (iv) *mean nacelle orientation*, used as a proxy for the incoming wind direction $\bar{\phi}_{10\,\mathrm{min}}$, and (v) *mean blade pitch angle*, $\bar{\beta}_{10\,\mathrm{min}}$.

The SCADA-TI is then derived as:

$$\mathrm{TI}_{10\mathrm{min}} = \frac{\sigma_{u,10\mathrm{min}}}{\overline{u}_{10\mathrm{min}}}. \tag{1}$$

2.2 LiDAR datasets

In addition to the SCADA measurements, profiling LiDAR measurements were made available at each wind farm to provide hub-height wind speed and direction timeseries. The measurements origin from Leosphere WindCubes, installed on the roof of the offshore substations (OSS) of each wind farm, each scanning across 12 vertical levels (up to $\approx 320\,\mathrm{m}$). The LiDAR positions are indicated in Fig. 1 as red triangles. For each level, the mean and standard deviation of the *horizontal wind speed*, $\bar{u}_{H,10\,\mathrm{min}}$ and $\sigma_{u_H,10\,\mathrm{min}}$ along with the *wind direction*, $\bar{\phi}_{10\,\mathrm{min}}$ are retrieved as 10-minute aggregated data. Their value at hub height is then estimated based on a simple linear interpolation while $\mathrm{TI}_{10\mathrm{min}}$ is computed based on Eq. 1 but using u_H as a proxy for the streamwise wind speed thus slightly diverging from the earlier definition.

At Westermost Rough, the LiDAR was installed on the OSS at $53°48'18''$N, $0°07'59''$E. The measurement campaign ran from 13 January 2016 to 6 December 2017, with scarce data availability over the period (Fig. 4). The LiDAR being located inside the wind farm, all its measurements are waked.

At Anholt, the LiDAR was installed at $56°35'44''$N, $11°09'10''$E, on the OSS at the western edge of the array. The measurement campaign spans 1 January 2013 to 31 December 2014 with mostly continuous availability over the period. Due to its location, some of the LiDAR measurements are affected by the wind farm wake, especially for easterly winds.

2.3 WRF datasets

To complement the SCADA and LiDAR site measurements, this study also makes use of mesoscale data generated within the DTWO project (Fischereit et al., 2026; DTWO Consortium, 2024; Larsén and the DTWO Consortium, 2025) . Specifically, WRF simulations were performed over the North Sea for two periods: a two-year period covering 2016-2017 for Westermost Rough, and a simulation covering July 2013 to June 2014 for Anholt.

2.3.1 WRF setup

The WRF setup consists of three nested domains with horizontal grid resolutions of 9, 3, and 1 km. Runs are executed in 8-day batches with a 1-day overlapping spin-up period. All neighboring wind farms commissioned during or before the simulation periods (*e.g.* Humber Gateway) are included using the Fitch parameterisation (Fitch et al., 2012), while Westermost Rough and Anholt themselves are not modeled. ERA5 reanalysis provides boundary conditions every 6 h, and spectral nudging is applied in the outermost domain to constrain atmospheric features with horizontal scales larger than 200 km. The other physical parameterizations used in the simulations are: short- and long-wave radiation following the RRTMG scheme (Iacono et al.,

2008), planetary boundary layer represented by the Mellor-Yamada-Nakanishi-Niino scheme (Nakanishi and Niino, 2006), surface layer using the Monin-Obukhov Janjić Eta scheme (Janjić, 1994), Noah (Chen and Dudhia, 2001) land surface model, convection following the Kain-Fritsch scheme (Kain, 2004), and microphysics using the WSM 5-class scheme (Hong et al., 2004).

2.3.2 WRF features

These WRF datasets are used to derive the atmospheric features and subsequently correlate wake models performance to the atmospheric state through these labels. Virtual probes are placed at the barycenters of both wind farms, to extract the local WRF values of the freestream *wind speed*, *wind direction*, *turbulence intensity*, and *stability class*; their hub-height wind velocity fields are also extracted. While SCADA data are available at a 10-minute resolution, WRF simulations were exported at a 30-minute resolution. All results reported here are subsequently aligned (*i.e.* no averaging) to the 30-minute timestamps and processed accordingly.

Atmospheric shear and veer are neglected: the flow is assumed horizontally homogeneous at the rotor scale. Stability classes are inferred from the inverse Monin-Obukhov length ($\mathrm{RMOL} = 1/L$) following the classification of surface layer stability provided in Table 1.

Stability class	**Abbreviation**	**Range of L [m]**
Very Stable	VS	$0 \leq L < 50$
Stable	S	$50 \leq L < 200$
Near-Neutral Stable	NNS	$200 \leq L < 500$
Neutral	N	$\lvert L\rvert \geq 500$
Near-Neutral Unstable	NNU	$-500 \leq L < -200$
Unstable	U	$-200 \leq L < -100$
Very Unstable	VU	$-100 \leq L < 0$

Table 1. Atmospheric stability classification based on the Monin-Obukhov length (L).

Complementary to the stability class, a time-varying TI is estimated from the WRF data following Larsén (2024), which proposes to use the Kaimal turbulence spectrum to approximate the wind speed component standard deviations, σ_u, σ_v, and σ_w, at given heights and wind speeds. Empirical ratios between σ_u and the other components then allow the turbulence kinetic energy (TKE) extracted from WRF to be converted into the along-wind standard deviation. The ambient TI is estimated as

$$\mathrm{TI} = \frac{\sigma_u}{\overline{u}} = \frac{1}{\overline{u}}\sqrt{\frac{2\,\mathrm{TKE}}{1+\alpha^{-1}+\beta^{-1}}}, \tag{2}$$

with $\overline{u}$ the time-averaged hub-height wind speed (consistent with u defined in Section 3.2) and α and β empirical, height- and wind-speed-dependent coefficients determined from Kaimal turbulence spectra.

Finally, flow heterogeneity is quantified by two velocity gradient indicators: an along-wind gradient $\alpha_{\parallel}$ and a cross-wind gradient $\alpha_{\perp}$. Both are estimated at each time step by fitting a plane to the hub-height velocity field sampled at the WRF grid points located within the farm footprint. Denoting by u_i and u_c the hub-height wind speed at the ith WRF grid point and at the barycenter of the wind farm respectively, the plane coefficients (α_x, α_y) are obtained by least-squares minimisation of the wind speed residuals:

$$\mathcal{R}(\alpha_x, \alpha_y) = \sum_{i=1}^{N_{\mathrm{gp}}} \left[u_i - (u_c + \alpha_x (x_i - x_c) + \alpha_y (y_i - y_c))\right]^2, \tag{3}$$

where N_{gp} is the number of WRF grid points within the farm polygon and $(x_c,\ y_c)$ is the position of the barycenter of the wind farm. The along- and cross-wind gradients are then obtained by projecting (α_x, α_y) onto the WRF-derived freestream wind direction ϕ at the same timestamp:

$$\alpha_{\parallel} = -\alpha_x \sin\phi - \alpha_y \cos\phi, \qquad \alpha_{\perp} = \alpha_x \cos\phi - \alpha_y \sin\phi. \tag{4}$$

Residuals are tracked to ensure the quality of the fit. Homogeneous inflow corresponds to $\alpha_{\parallel} = \alpha_{\perp} = 0$, whereas deviations from zero reflect the presence of mesoscale gradients.

3 Methodology

The different sources of uncertainties presented in the introduction are assessed by running the workflow under different configurations (*e.g.* with WRF or SCADA-filtered inflow data, or across FLORIS, PyWake, and FOXES) and comparing the resulting power estimates to quantify variability across workflows.

The following subsections describe the SCADA-filtering and wake modelling stages of the workflow in detail.

3.1 SCADA filtering

The first stage of the workflow consists of filtering the raw SCADA data to retrieve filtered turbine power production as reference data for model validation. In addition, *ambient wind speed* u, *wind direction* ϕ and *Turbulence Intensity* TI timeseries are derived directly from the nacelle measurements. The ambient timeseries are intended to represent the freestream inflow conditions at the wind farm for each timestamp, and are used both as model input and for binning the results.

No standardised process exists to derive the inflow states from the SCADA data. The approach in this paper is to derive them independently at two of the contributing institutes: at Fraunhofer IWES with an internal toolchain (Vollmer et al., 2024), and at VKI with a workflow based on the open-source FLASC (Fleming et al., 2022) toolchain. This makes it possible to assess the portion of uncertainty in the validation reference and the inflow wind that arises solely from differences in raw data processing.

The process of filtering the power production involves identification of the turbine operating status using the available SCADA channels. One by-product of the filtering are nacelle-anemometer or SCADA power curves derived from the filtered power data. As the filtering process differs between the institutes, the resulting SCADA power curves can also slightly differ.

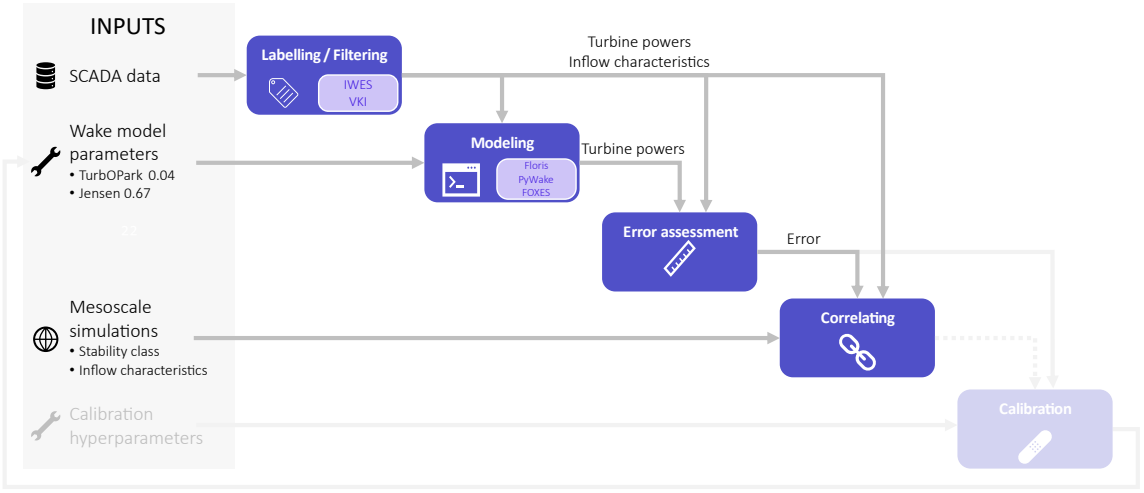


Figure 2. SCADA-driven wake model validation workflow. The WRA stage (labelling/filtering) can be substituted with WRF or LiDAR inflow data without modifying the downstream stages. The calibration block (greyed out) is shown for completeness but is outside the scope of this paper.

For deriving the ambient wind timeseries, the nacelle measurements of wind speed and wind direction undergo several quality controls and correction methods. The two SCADA-processing workflows are explained in more detail in the following sections.

3.1.1 Fraunhofer IWES workflow

The in-house workflow at Fraunhofer IWES has been developed in the context of commercial projects and has, for example, been used for model validation in the German X-Wakes project (Vollmer et al., 2024). It follows a three-step procedure designed to generate timeseries of corrected wind speed and direction at each turbine, as well as an estimate of the turbines' power production under 100 % availability. The three steps are:

1. flagging turbine performance that deviates from standard operating conditions;
2. correcting wind direction and wind speed during standard operation;
3. gap-filling all three parameters during periods flagged as having deviating performance.

Turbine underperformance is identified using variance filters and polygon filters in wind-speed-power and pitch-power space. Wind speed and wind direction measurements are flagged and filtered for implausible values and trends between consecutive timestamps. Flagged wind speeds are also flagged in the power signal for turbine underperformance.

Wind direction measurements are corrected for two main error sources, temporal inconsistency and incorrect north-reference calibration, using a FOXES wake simulation as a reference.

The gap-filling approach is closely aligned with the German TG10 guideline (FGW, 2024), which involves discarding all meteorological measurements during turbine downtimes and replacing them with corrected external data. In this workflow, 100 m wind speed from ERA5 is used as the external data source. A correction to the specific wind conditions at each turbine is applied by training a k-nearest-neighbours algorithm on periods of standard operation, using wind direction and month of the year as input features.

Freestream conditions are obtained by averaging the previously corrected wind direction over all turbines, and the wind speed over only those turbines classified as free from wake effects. The classification is based on a wake model run for each wind direction, which identifies turbines unaffected by internal wakes and therefore suitable as freestream reference points.

3.1.2 VKI workflow

The in-house workflow at VKI is based on the open-source FLASC (Fleming et al., 2022) toolchain and aims to generate a consistent, filtered SCADA dataset together with representative freestream inflow conditions. It follows a four-step procedure:

1. preprocessing and filtering of raw SCADA data;
2. correction of wind direction measurements (northing calibration);
3. consistency filtering across turbines;
4. determination of freestream conditions from upstream turbines.

Raw SCADA data are first converted into the FLASC internal format and subjected to a sequence of filtering steps. For each timestamp, measurements with missing values in wind speed or power are discarded, as are non-physical values (e.g. negative or unrealistically large wind speeds). Data are further filtered based on turbine operating conditions, including minimum power thresholds and curtailed operation. Sensor-stuck faults are identified using a standard-deviation threshold applied over three consecutive measurements. Finally, outliers in the power curve are removed using an iterative, median-based filtering approach. No gap-filling is performed, and filtered data points are excluded from further analysis.

Wind direction measurements are subsequently corrected through a multi-step northing calibration procedure. First, the consistency of directional offsets between turbines is evaluated to detect temporal discontinuities. A homogenization with the HOGER method (NatLabRockies, 2026; Fleming et al., 2022), developed within the TWAIN project, is then applied to remove relative temporal variations in turbine northing calibration. For a single turbine, the residual steady offset is corrected by aligning the energy-ratio curve derived from the filtered SCADA data with the corresponding curve obtained from FLORIS wake model simulations, yielding a reference wind direction. Subsequently, the remaining turbine-specific biases are estimated using this reference as an initial guess and corrected through alignment of energy-ratio curves. Additional consistency checks account for inter-turbine dependencies: measurements from a given turbine are flagged as invalid if upstream turbines influencing its operation are themselves flagged as faulty.

Freestream conditions are derived using a wake-model-based classification of upstream turbines. FLORIS simulations are performed over the full wind rose using an enlarged wake expansion coefficient to conservatively identify turbines operating in

unwaked flow. For each timestamp, wind direction is first averaged across all turbines to obtain a farm-level reference direction, after which the upstream wind direction, wind speed and TI are computed by averaging only the turbines classified as upstream.

The power curves derived from the VKI-filtering workflow are shown in Fig. 3 for both sites.

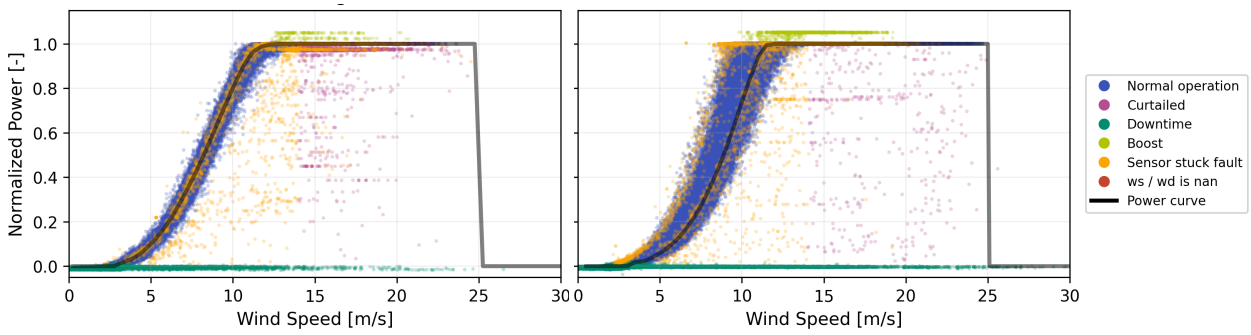


Figure 3. SCADA-derived power curves from the VKI-filtering workflow: **(Left)** Westermost Rough. **(Right)** Anholt.

3.2 Wake modelling

The freestream inflow is parametrised in terms of a single freestream hub-height wind speed, u, wind direction, ϕ, and turbulence intensity, TI, per timestamp. The turbine power curves are constructed from the filtered SCADA data while the thrust curves are retrieved from publicly available data sources (Bulder et al., 2016; Lanzilao and Meyers, 2021; wind-turbine-models.com, 2026; commissie mer, 2021). Wind farm simulations are performed using the open-source toolboxes FLORIS (NREL, 2025), PyWake (Pedersen et al., 2023) and FOXES (Schmidt et al., 2023) with the Wind Farm API (WIFA) (Quick et al., 2024, 2025) ensuring consistent input data across the three software packages. The wake models are run in a quasi-steady configuration and ignoring turbine downtimes and other abnormal operations: all turbines are considered as running according to their power curve. While WRF velocity field would allow to drive the wake model run using a heterogeneous inflow, the inflow derived from the SCADA data does not. All simulations are therefore performed using an homogeneous inflow configuration for a fairer inter-inflow comparison. Two wake models are considered: Jensen (Jensen, 1983) and TurbOPark (Nygaard et al., 2020).

Perhaps as the most classical engineering wake model, Jensen (Jensen, 1983) assumes that the wake expands as a cone with a fixed linear expansion coefficient k_{wk}. In this study, the original Jensen model is adapted to account for the effect of TI on the wake recovery following:

$$k_{\mathrm{wk}} = k_{\mathrm{wk,TI}}\,\mathrm{TI} \tag{5}$$

where $k_{\mathrm{wk,TI}} = 0.67$ is a calibration parameter as proposed by EMD International A/S (2025). For a standard offshore TI of 6%, this corresponds to $k_{\mathrm{wk}} = 0.04$, which is the default value used for offshore conditions in the original Jensen model. Wakes are superimposed quadratically in the simulations with the Jensen model, using the freestream velocity as the reference (Katic et al., 1987).

The TurbOPark model (Pedersen et al., 2022) was specifically developed to mitigate excessive wake recovery over long downstream distances through a modified turbulence recovery scheme. The model uses a non-linear wake expansion sensitive to the local TI and its downstream decay:

$$\frac{\mathrm{d}\sigma_{\mathrm{wk}}(x)}{\mathrm{d}x} = A\,\mathrm{TI}_{\mathrm{eff}}(x) = A\sqrt{\mathrm{TI}^2 + \mathrm{TI}^2_{\mathrm{wk}}(x)}, \tag{6}$$

where σ_{wk} denotes the characteristic wake width and A is a calibration parameter. The effective local turbulence intensity, $\mathrm{TI}_{\mathrm{eff}}(x)$, at downstream distance x is obtained as the quadratic sum of the ambient turbulence intensity TI and the wake-added turbulence intensity $\mathrm{TI}_{\mathrm{wk}}(x)$, whose evolution is modelled following Frandsen (2007). The TurbOPark model is run with the default expansion coefficient $A = 0.04$ and with quadratic superposition of wake deficits, using the freestream velocity as the reference.

Both models are fed with a basic set of model parameters defining the wake model through the WIFA interface. Most notably, rotor averaging is performed on a 3×3 grid, centered at the turbine hub with no correction for shear or veer. The complete set of parameters used is reported in Appendix A. In this study, it is not further investigated how these parameters are interpreted internally by each software package, and which default settings are used for parameters not explicitly specified. The model outputs are therefore expected to differ both from each other and from the canonical formulations in the original model references. This residual variability is part of the inter-software uncertainty assessed in Sect. 5.2.

3.3 Metrics

Three performance metrics are used throughout this study: they are first applied to the cross comparison of the inflow state inferred from different data sources and then to the assessment of the wake model power estimates with reference to the SCADA data. The performance metrics considered are the *bias*, b, the *root mean square error* (RMSE), ε, and the *Pearson correlation coefficient*, r. For power, b, ε, and r are computed using power normalized by the rated power, $P^* = P/P_{\mathrm{rated}}$ [-]. For inflow estimation, the same metrics are computed from the differences between the estimated and reference inflow variables. Metrics are evaluated at the wind farm-level by first computing them for each turbine and then averaging across all turbines.

4 Analysis of the different inflow sources

This section discusses the inflow characterisation at both sites. The atmospheric features derived from the WRF simulations are first presented in Section 4.2, providing insight into the local flow physics at Westermost Rough and Anholt. The consistency of these WRF-derived features with the SCADA and LiDAR field measurements is then assessed in Section 4.3.

4.1 Data availability

The monthly data availability for both sites is summarised in Fig. 4 (bottom). For the SCADA data, data availability is defined as the fraction of turbines from which exploitable power measurements were retrieved after filtering for any given timestamp.

For the WRF and LiDAR data, data availability at a given timestamp is either 100% or 0%, depending on whether valid wind speed and direction measurements are available. All data availability metrics are averaged over one-month windows.

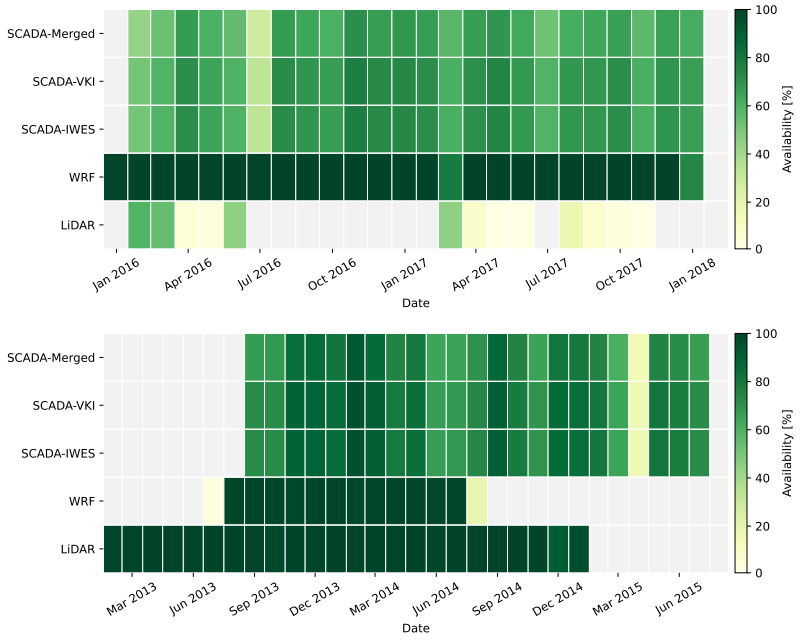


Figure 4. Data availability: monthly fraction of turbines with exploitable SCADA measurements, and binary availability of WRF and LiDAR data. **(Top)** Westermost Rough. **(Bottom)** Anholt.

4.2 WRF-derived features

The WRF labels presented in Section 2.3 allow to better understand the wind climatology at both site and build insight from about large scale trends, unavailable to the punctual SCADA measurements. These labels are shown in Fig. 5 where the wind rose, stability rose and wind speed gradients directional distributions are reported.

At Westermost Rough (Fig. 5, top row), winds are predominantly from the south-west with a sharp peak around 250° (West-South-West). Unstable conditions dominate the stability rose while stable conditions occur predominantly for winds with an onshore component. Both the mean cross-wind gradient $\alpha_{\perp}$ and the along-wind gradient $\alpha_{\parallel}$ remain small in the easterly sector, consistent with the long and homogeneous sea fetch in that direction. By contrast, larger mean gradient magnitudes are observed in the westerly sector, indicative of the upstream presence of land and of the flow adjusting to the reduced surface roughness of the offshore environment. The wake signature of the neighbouring Humber Gateway wind farm is also clearly visible around 165° (South-South-East) as a sharp peak in cross-wind velocity gradient.

Similar observations can be made from the directional distributions at Anholt (Fig. 5, bottom row). Winds blow predominantly from a south-easterly to north-westerly sector with unstable conditions dominating the stability rose. Fewer homoge-

neous conditions are encountered than at Westermost Rough, probably as a result of the larger spatial extent of the Anholt farm. Onshore sectors are nonetheless associated with overall higher gradients, while a significant dip in cross-wind gradient is observed to the north-east, in the direction of Anholt island.

Figure 6 finally reports the joint distribution of the WRF-derived ambient TI and stability class at both farms. A near-linear dependence is visible at both sites, with TI decreasing from approximately 10-12 % in very unstable conditions to 3-4 % in very stable conditions. The bin-mean values are $\mathrm{TI} = 7.3\,\%$ at Westermost Rough and $8.8\,\%$ at Anholt. The distribution at Westermost Rough is bimodal and somewhat more spread across stability classes than at Anholt where the bulk of the data is concentrated around near-neutral conditions. This is consistent with the literature where mean TI values in the range of 5-8% are frequently reported (*e.g.* Larsén et al. (2026); Hansen et al. (2012)). The default value $\mathrm{TI} = 6$ % often considered as representative of offshore conditions for engineering wake-model studies (*e.g.* Doekemeijer et al. (2022); Van Binsbergen et al. (2024)) therefore lies below the timeseries-mean at both sites. Figure 6 nevertheless shows that it corresponds approximately to the TI associated with neutral conditions. While selecting this default value, $\mathrm{TI} = 6$ % allows the atmospheric-related uncertainty to be assessed in an unbiased way with respect to stability, it might consequently introduce a small negative bias from an AEP perspective.

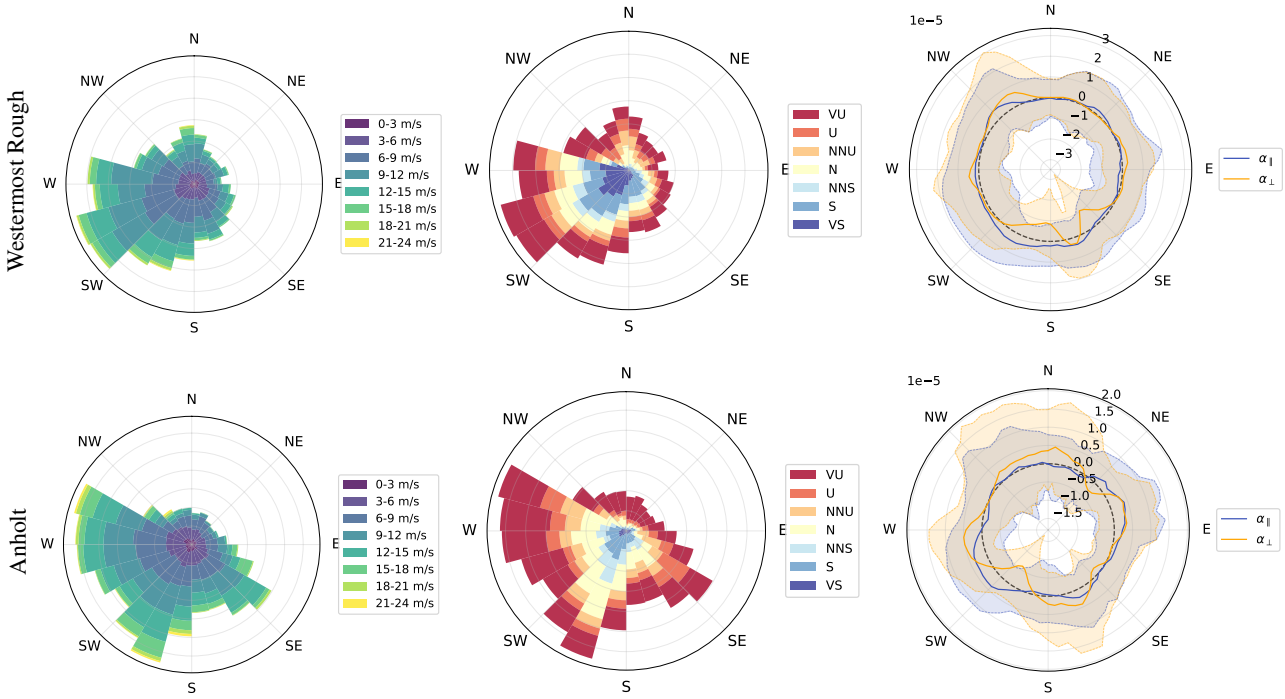


Figure 5. WRF-derived directional distributions: **(Left)** wind speed rose. **(Center)** stability rose. **(Right)** along- and cross-wind gradient components: mean value (full line) and P10 and P90 (shaded area). **(Top)** Westermost Rough. **(Bottom)** Anholt.

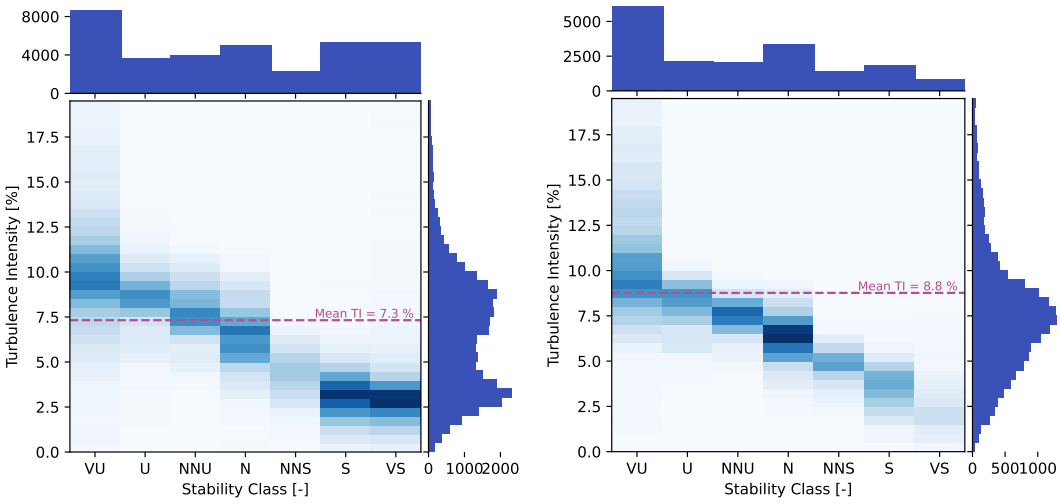


Figure 6. Joint distribution of turbulence intensity and stability class for **(Left)** Westermost Rough and **(Right)** Anholt. Marginal distributions: stability class **(Bottom)** and TI **(Right)**. Dashed red line indicates the mean TI.

4.3 Comparison of derived features across inflow sources

As discussed previously, the definition of the inflow remains ambiguous, with each WRA methodology suffering from its own limitations. In an attempt to mitigate workflow-dependent uncertainties, we assess the variability across the different inflow sources available, namely *SCADA-VKI*, *SCADA-IWES*, *LiDAR*, and *WRF*. This comparison is shown in Figs. 7 and 8 for the two sites and for wind speed and wind direction, respectively. The upper part of these plots provides pairwise performance matrices for each metric, while the lower part shows the corresponding pairwise correlation scatter plots. Prior to comparison, all timeseries are temporally aligned on the WRF timestamps. For the LiDAR data, timeseries are split between *waked* and *unwaked* timestamps. For Anholt, all LiDAR timestamps with a wind direction within $[180°, 340°]$ are flagged as unwaked while, at Westermost Rough, all LiDAR timestamps are marked as waked since the LiDAR is located inside the farm. For a fair comparison, waked conditions are shown in the correlation scatter plots but are excluded from the performance matrices of Anholt. Still, for Westermost Rough, the performance matrices are computed including waked conditions as discarding them would essentially remove all available data points from the comparison.

We computed b , ε and r for each pair of sources, with the x-axis source as reference, *i.e.* positive b indicates that the y-axis source overestimates the x-axis source. Better performance corresponds to lighter colour shades in the matrices.

Considering wind speed and direction, the scatter plots (Figs. 7 and 8, bottom panels) reveal distinct behaviours depending on which inflow pairs are compared. The SCADA-SCADA panels collapse onto the 1:1 line at both farms for both the wind speed and wind direction meaning the two filtering workflows produce nearly interchangeable inflow timeseries. For the LiDAR, the wind direction based filtering creates a sharp cutoff between waked and unwaked condition in the wind direction correlation plot. Simultaneously, the transition from waked to unwaked conditions appears as a diffuse shift in the wind speed scatters.

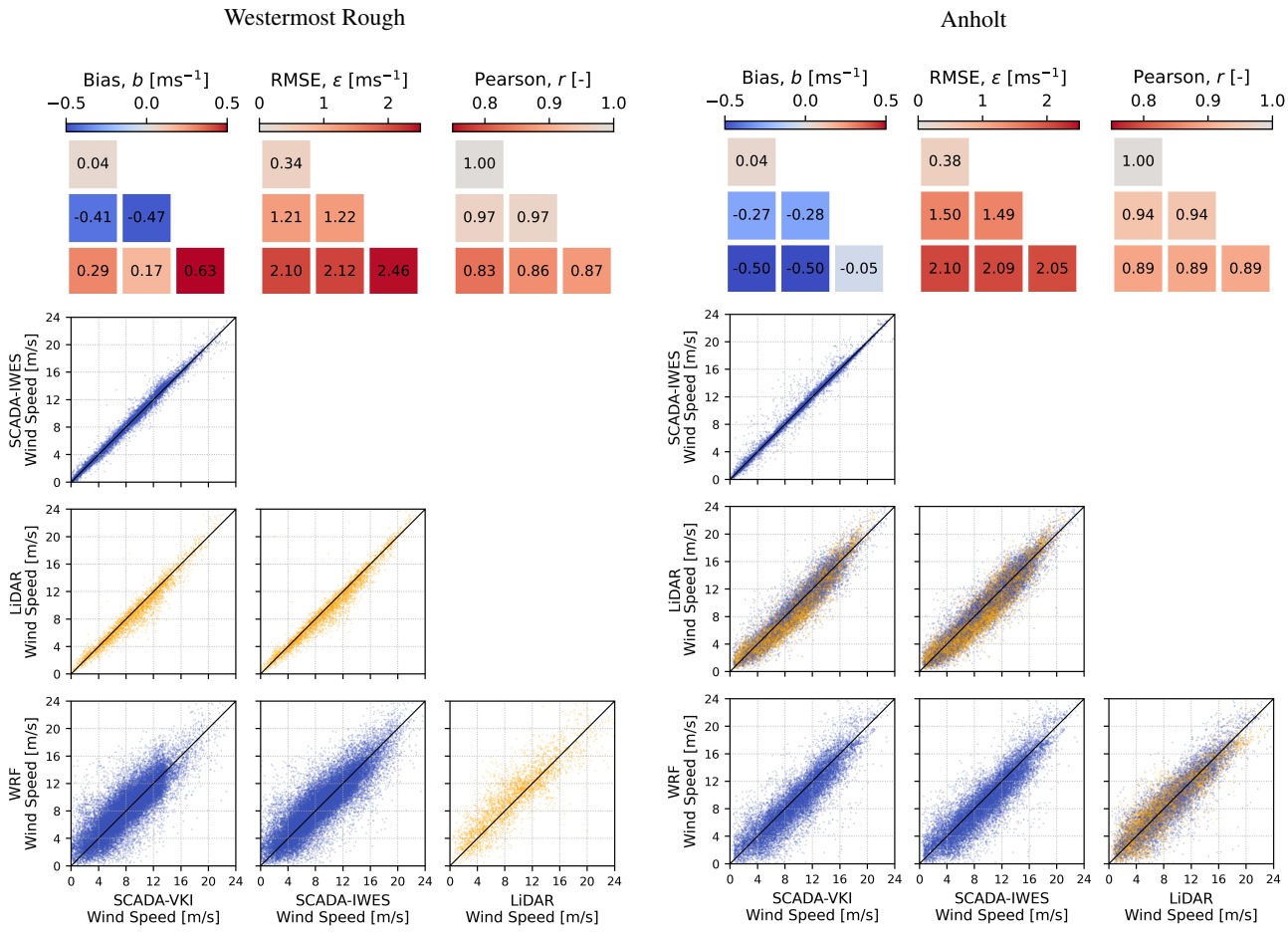


Figure 7. Wind speed performance matrices and correlation plots for **(Left)** Westermost Rough and **(Right)** Anholt. The cells of the performance matrices follow the same ordering as the correlation plots. **(Top per panel)** Pairwise performance matrices: b [m s^{-1}], ε [m s^{-1}], and r [-] across all inflow sources. Each cell shows the metric for the y-source relative to the x-source. **(Bottom per panel)** Pairwise correlation scatter plot; blue: unwaked conditions; orange: waked conditions.

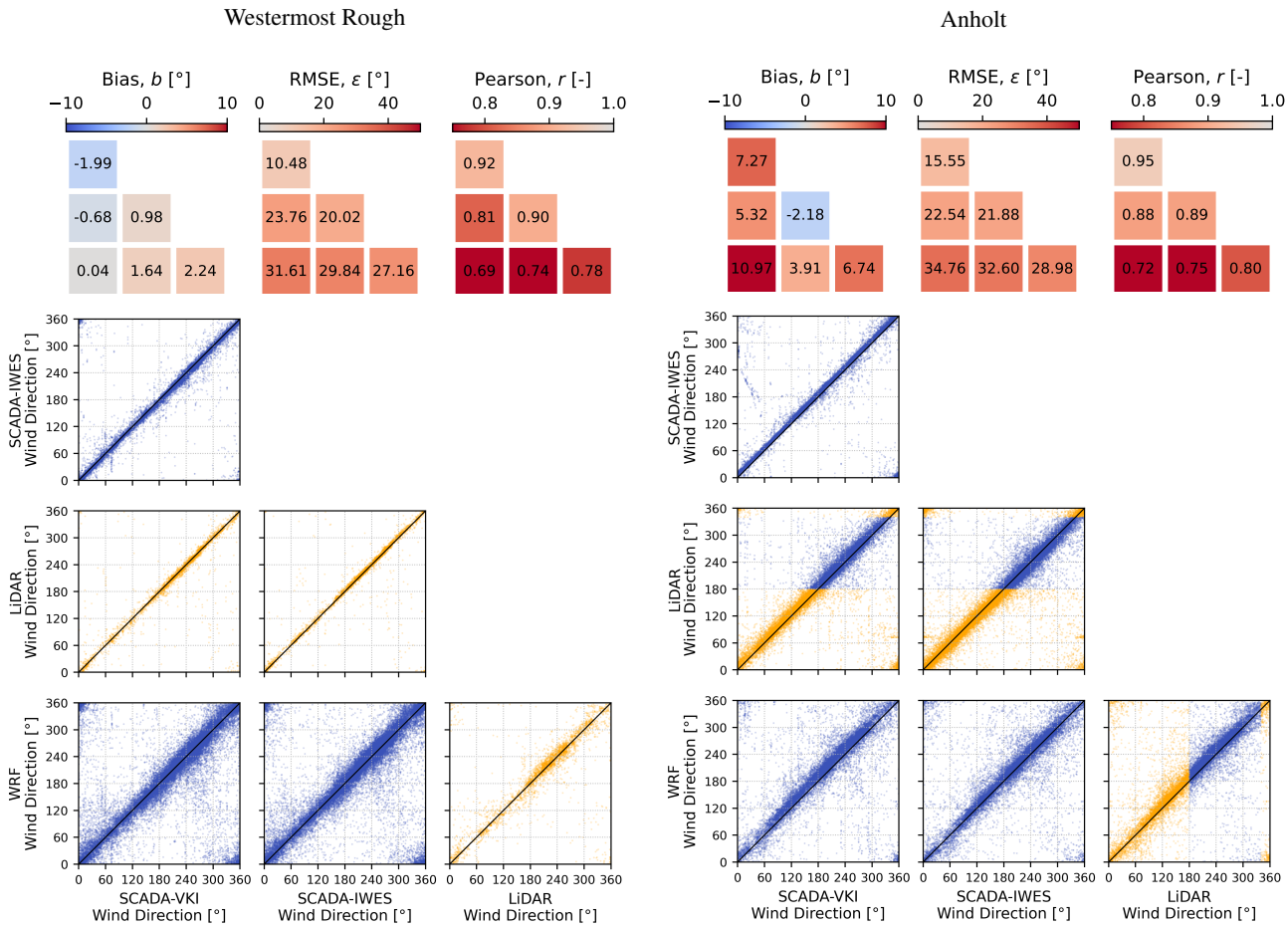


Figure 8. Wind direction performance matrices and correlation plots for **(Left)** Westermost Rough and **(Right)** Anholt. The cells of the performance matrices follow the same ordering as the correlation plots. **(Top per panel)** Pairwise performance matrices: b [°], ε [°], and r [-] across all inflow sources. Each cell shows the metric for the y-source relative to the x-source. **(Bottom per panel)** Pairwise correlation scatter plot; blue: unwaked conditions; orange: waked conditions.

Wind speeds are underestimated by the LiDAR for waked conditions at below-rated speeds and converge back toward the 1:1 line from rated. Once waked timestamps are excluded, the scatter recenters onto the diagonal. The WRF panels show a broader cloud at both sites, indicating a substantial loss of temporal correspondence, though the cloud remains centered on the 1:1 line.

The wind speed performance matrices (Fig. 7) quantitatively confirm these observations. The SCADA intercomparison (upper triangle) shows excellent agreement. Inter-workflow biases are $0.04\,\mathrm{m\,s^{-1}}$ at Westermost Rough and Anholt, with RMSE values of 0.34 and $0.38\,\mathrm{m\,s^{-1}}$ respectively and near-perfect correlation ($r \geq 0.99$). These biases are all the more small when compared to typical nacelle anemometer measurement uncertainties of order $0.1\,\mathrm{m\,s^{-1}}$ (Beltrán et al., 2013). For the LiDAR, the wind speed bias reaches $-0.47\,\mathrm{m\,s^{-1}}$ at Westermost Rough (where waked timestamps are retained) and $-0.28\,\mathrm{m\,s^{-1}}$ at Anholt (where they are excluded), with correlations remaining high ($r \geq 0.94$). The WRF row shows that the wind speed bias flips sign between sites: WRF overestimates at Westermost Rough (up to $+0.29\,\mathrm{m\,s^{-1}}$) and underestimates at Anholt (down to $-0.50\,\mathrm{m\,s^{-1}}$) compared to the SCADA data. WRF's RMSE is essentially independent of the reference ($\varepsilon \approx 2.05$-$2.12\,\mathrm{m\,s^{-1}}$ at both farms compared to the SCADA) when excluding waked conditions. Finally, slightly degraded wind speed correlations are noted, $r \approx 0.85$-0.89, when comparing WRF to the SCADA sources.

For wind direction (Fig. 8), the SCADA intercomparison reveals a small bias of $-1.99°$ at Westermost Rough (RMSE $10.48°$, $r = 0.92$) whereas a much larger bias of $7.27°$ is reported at Anholt (RMSE $15.55°$, $r = 0.95$). This discrepancy at Anholt likely reflects differences in per-turbine northing correction between the two workflows. The LiDAR wind direction biases follow a similar trend with lower biases observed at Westermost Rough ($|b| \leq 0.98°$). Despite its waked conditions, the LiDAR is able to capture the wind direction with good agreement with the SCADA data. At Anholt, the bias between the LiDAR and the SCADA-IWES wind direction is smaller ($b = -2.18°$) than between the LiDAR and the SCADA-VKI data ($b = 5.32°$). The WRF wind direction bias remains small at Westermost Rough ($|b| \leq 2.24°$) but reaches almost $11°$ at Anholt while RMSE stays close to $30°$ at both sites, and correlations ranges between 0.69 and 0.80.

The TI comparison is less conclusive, primarily because no unambiguous TI baseline can be defined. Each source resolves turbulence over a different averaging volume and with a different sensitivity to local flow features: the nacelle anemometers sit downstream of the rotor, the profiling LiDAR are subject to probe-volume averaging and cross-contamination, and the WRF does not resolve fine-scale turbulence. A rigorous cross-source assessment would therefore require dedicated corrections for each instrument, which is beyond the scope of this work. In the absence of a true reference, and since the LiDAR-WRF agreement becomes reasonable once waked conditions are excluded, the WRF-derived TI is retained as the reference in the following. The pairwise correlation plots and performance matrices supporting this choice are reported in Appendix B.

Overall, Figures 7 and 8 support the view that, even though WRF lacks the temporal accuracy of the SCADA-filtered datasets, it provides a reasonable estimate of the mean wind speed and direction and can be used as a proxy for large-scale flow features. Only two inflow sources are retained for the wake modelling results of Section 5: a *SCADA-Merged*, dataset, which provides the best temporal accuracy, and a WRF dataset. The *SCADA-Merged* dataset, hereafter referred to as *SCADA* dataset, is constructed by averaging the SCADA-VKI and SCADA-IWES on a time step basis. However, only the wind speed and wind direction are directly estimated based on the SCADA-data. Due to the high sensitivity of wake model to the inflow TI and the large variability of the TI estimates across sources, the WRF-derived TI is indeed retained for both inflow source

thus overriding the SCADA-derived TI. A similar approach was notably used by Mayol et al. (2026) who used synthetic WRF TI measurements for WRA.

5 Wake modelling results

The wake modelling framework introduced in Section 3.2 is now applied to both wind farms. The resulting power timeseries are compared against the SCADA reference. This comparison addresses three key questions: How does the workflow affect the wake model performance metrics? To what extent does the atmospheric state influence the accuracy of the wake model results? And how do these discrepancies appear in the temporal evolution of the power timeseries across seasonal and diurnal timescales?

5.1 Baseline reference power

The operational flags considered are *normal operation*, *curtailed*, *power-boost*, *downtime*, and *invalid*. Each SCADA-filtering workflow may however use a different methodology to derive turbine operational flags. As a result, one workflow may classify a given timestamp as *downtime*, while another may classify the same timestamp as *invalid*. This can lead to different sets of valid timestamps and, consequently, different reference power timeseries.

After verifying the consistency of the turbine operational flags, the outputs of both workflows are merged to define a common set of valid timestamps. For a given timestamp, if the two SCADA workflows agree on the turbine's operational status, the corresponding power value is retained in the reference timeseries. If the workflows disagree, the timestamp is flagged as invalid. This merged reference power timeseries is referred to as SCADA.

Timestamps associated with widespread turbine unavailability or operational disruptions within the windfarm could significantly skew the error-metric assessment if included as-is. Two filtering steps are therefore applied to the merged reference power timeseries: at the turbine-level, any power measurement not classified as normal operation is discarded, i.e., set to NaN. At the farm-level, all turbine power measurements for a given timestamp are discarded if more than 75 % of them are simultaneously classified as neither normal operation nor power-boost condition, since they are unlikely to be representative of a standard wind farm operation. Power-boost measurements are nonetheless kept as boosting means small power increments without substantial effects on the turbine's thrust coefficient curve, and, as consequence, on wake effects between neighboring turbines.

5.2 Wind-farm performance metrics

The performance matrices are shown in Figs. 10 and 11 for Westermost Rough and Anholt, respectively. Comparable to the analysis of the inflow, b [-], ε [-], and r [-] are reported. Power is normalized by the rated power, $P^* = P/P_{\mathrm{rated}}$ [-] prior to any metric computation. For each cell, the metrics are computed at the turbine-level and then averaged over the wind farm, allowing each turbine to weight uniformly in the error, regardless of its data availability. The first column, framed in black, compares

each wake model run to the SCADA reference power while the remaining cells correspond to inter-model comparisons. Results for Jensen and TurbOPark are reported separately.

5.2.1 Comparison of modelled power against the SCADA reference

The comparison first focuses on the modelled power against the SCADA reference, i.e. on the framed column of the matrices. Figure 9 provides a qualitative comparison of the wake model powers against the reference SCADA power. The trends observed are the same across the two sites. The SCADA-derived inflow leads to better results than WRF, an observation then validated by the performance matrices in Figs. 10 and 11. Biases remain very small thanks to farm-level averaging and to the filtering applied by the power curve. Correlations remain above 0.93 with respect to the SCADA reference, slightly lower than for the inflow itself.

The WRF-driven results follow the same general trend but exhibit higher RMSE and lower correlation. Interestingly, the bias is not consistently better for the SCADA inflow: WRF-driven results even show slightly improved bias at Westermost Rough for the TurbOPark model, nearing zero bias. However, even for this specific case, concluding that WRF outperforms SCADA would be misleading since the RMSE and correlation are worse for WRF than for the SCADA inflow. This apparent bias improvement may stem from WRF's wind speed bias compensating for intrinsic wake model bias in specific cases. This observation strongly supports moving away from AEP-based validation toward time-series validation, as AEP-based approaches focus solely on mean total bias and thus neglect much of the problem's physical complexity.

5.2.2 Inter-model and inter-software comparison of the modelled power

Focusing on the inter-model comparison, little can be inferred from the scatter plots alone (Fig. 9): the Jensen and TurbOPark point clouds look similar. The performance matrices however reveal clear differences in the behaviour of the two models. The Jensen model exhibits homogeneous performance across software implementations with a maximum variability of 0.01 across software packages, confirming the consistency of this simple model across toolboxes. This is not the case for TurbOPark whose more complex implementation results in substantial inter-software variability in the metrics despite all inputs through WindIO being equal. For instance, at Anholt, the bias spans -0.02 to $+0.02$, with respect to the SCADA, across the three software implementations. Being more complex than Jensen, TurbOPark requires a larger number of implementation choices, each of which may influence the final result. Differences may therefore emerge from elements such as the details of wake superposition or rotor-averaging algorithms. The correlation and RMSE are however more uniform across software and inflow choices. Jensen appears to slightly outperform TurbOPark when investigating the RMSE and Pearson correlation, but these differences are too small to be really conclusive.

These observations additionally highlight an important methodological point: while accurate WRA is a necessary step toward reliable yield assessment, the uncertainty arising from the wake modelling procedure, down to the choice of software implementation, can be nearly as impactful as the WRA choice itself. With identical inputs and identical models, different software implementations yield noticeably different results. This calls for caution when reusing wake model parameters calibrated in a different software environment, as the optimal values may not necessarily transfer.

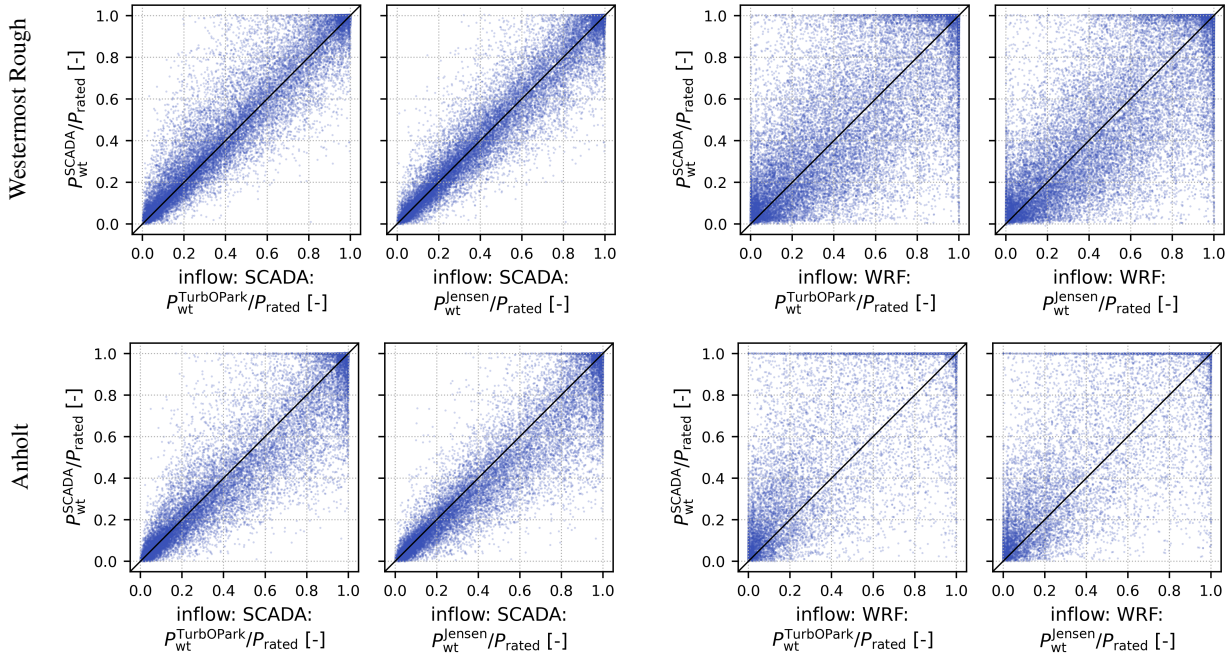


Figure 9. Power pairwise correlation scatter plots, SCADA against wake model (Jensen and TurbOPark, averaged over all software packages): **(Top)** Westermost Rough. **(Bottom)** Anholt. **(Left)** SCADA-derived inflow **(Right)** WRF-derived inflow. The solid line marks the 1:1 line.

5.3 Impact of atmospheric conditions

Figures 12 and 13 compare the SCADA reference power to the mean across all software implementations and wake models for different atmospheric states. For the sake of conciseness, only the SCADA inflow source is analysed here. Since no strong overperformer was identified in the previous section, the results for all software implementations and wake models are averaged together.

The matrices are constructed in three steps. First, the data are binned according to the chosen pair of atmospheric labels. Each bin, denoted $\mathcal{B}_i$, consists of a set of $N_{\mathcal{B}_i}$ timestamps sharing the same pair of atmospheric labels. Wind speed, u, and wind direction, ϕ, are taken directly from the SCADA-derived inflow, whereas quantities not directly retrievable from field measurements, namely the atmospheric stability class and the wind speed gradient magnitude $|\alpha| = \sqrt{\alpha_{\parallel}^2 + \alpha_{\perp}^2}$ (defined in Eqs. 3-4), are recovered from the WRF datasets. It should be reminded that the WRF-derived labels are proxies for the actual atmospheric state rather than direct measurements and that some timestamps may therefore be misclassified when binned based on this data.

Within each bin, the mean error is first computed individually for each turbine wt:

$$e_{\mathrm{wt}}(\mathcal{B}_i) = \frac{1}{N_{\mathcal{B}_i}} \sum_{t \in \mathcal{B}_i} \left(P_{\mathrm{wt}}^{\mathrm{wm}}(t) - P_{\mathrm{wt}}^{\mathrm{SCADA}}(t) \right) = \frac{1}{N_{\mathcal{B}_i}} \sum_{t \in \mathcal{B}_i} \left(\Delta P_{\mathrm{wt}}(t) \right), \tag{7}$$

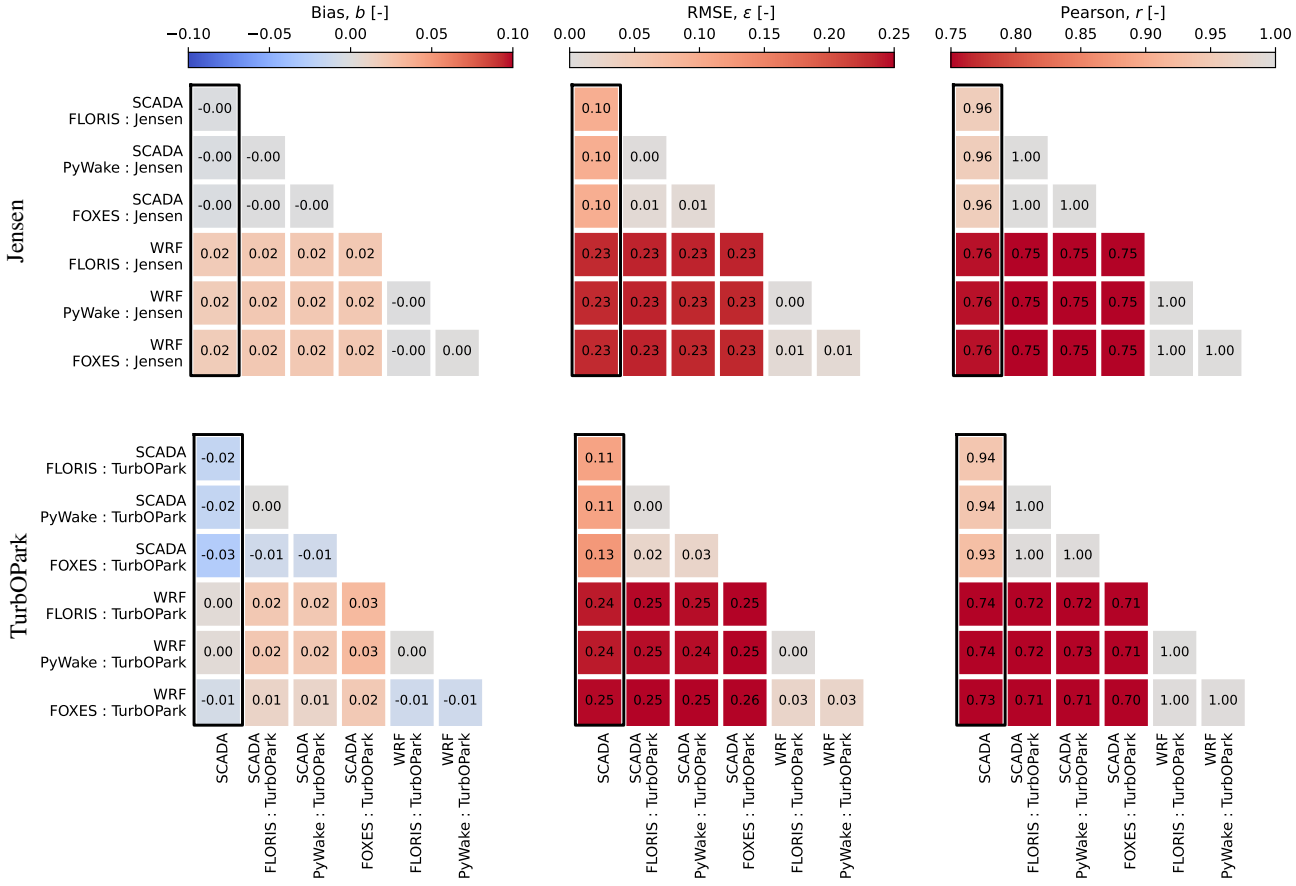


Figure 10. Westermost Rough, power performance metrics: **(Top)** Jensen model. **(Bottom)** TurbOPark model. Power is normalised as $P^* = P/P_{\text{rated}}$ [-]. Bias b [-], RMSE ε [-], and Pearson correlation coefficient r [-] are reported for each pair of inflow sources and software implementations.

where $P_{\text{wt}}^{\text{wm}}(t)$ and $P_{\text{wt}}^{\text{SCADA}}(t)$ denote the normalized wake-modelled and reference power at turbine wt and timestamp t, respectively. As discussed in Section 5.1, only timestamps and turbines classified under normal operation are retained at this stage. The per-turbine errors are then aggregated over the N_{wt} turbines of the farm:

$$E_{\text{mean}}(\mathcal{B}_i) = \frac{1}{N_{\text{wt}}} \sum_{\text{wt}=1}^{N_{\text{wt}}} e_{\text{wt}}(\mathcal{B}_i), \qquad \text{and} \quad E_{\text{MAE}}(\mathcal{B}_i) = \frac{1}{N_{\text{wt}}} \sum_{\text{wt}=1}^{N_{\text{wt}}} |e_{\text{wt}}(\mathcal{B}_i)|, \tag{8}$$

so that all turbines contribute equally to the aggregated metric: $\frac{1}{N_{\mathcal{B}}} \sum_i E_{\text{mean}}(\mathcal{B}_i) = b$/ Now considering a bin, $\mathcal{B}_{\text{all}}$, covering all samples available, the time-averaged farm-bias is recovered so that $E_{\text{mean}}(\mathcal{B}_{\text{all}}) = b$ while $e_{\text{wt}}(\mathcal{B}_{\text{all}})$ is simply the time-averaged bias of turbine wt.

The resulting binned power errors are shown in Figs. 12 and 13. The farm mean E_{mean} (top panels) provides an indication of the total wind farm error within the selected bin (*i.e.* bias). Since averaging errors across the wind farm causes bias compen-

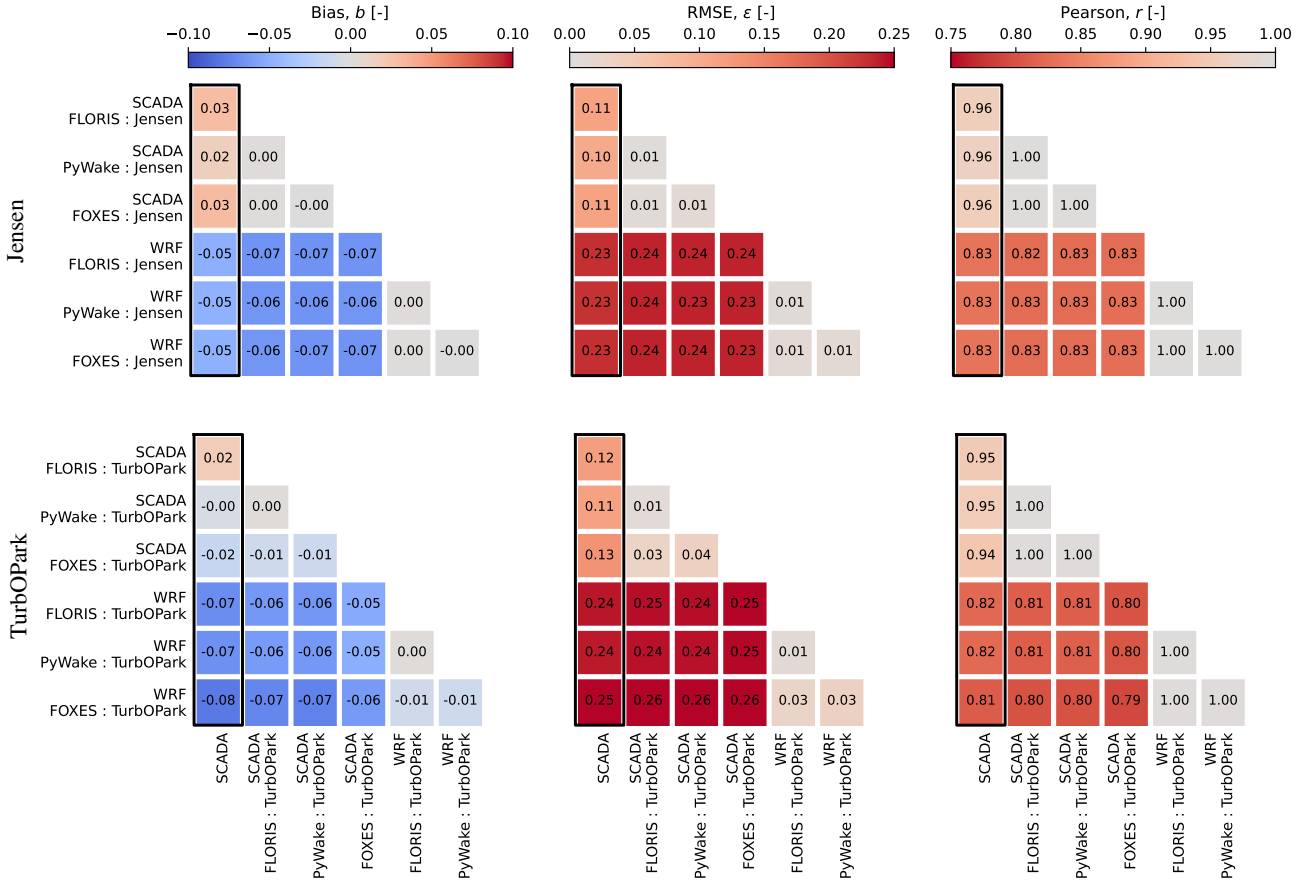


Figure 11. Anholt, power performance metrics: **(Top)** Jensen model. **(Bottom)** TurbOPark model. Power is normalised as $P^* = P/P_{\text{rated}}$ [-]. Bias b [-], RMSE ε [-], and Pearson correlation coefficient r [-] are reported for each pair of inflow sources and software implementations.

sation between turbines, the farm-Mean Absolute Error (MAE) E_{MAE} (middle panels) is added. Finally, the sample count $N_{\mathcal{B}_i}$ in each bin is reported in the bottom panels, and bins with $N_{\mathcal{B}_i} < 50$ are discarded to ensure sufficient statistical convergence.

Figures 12 and 13 reveal a more nuanced picture than that presented in Section 5.2. The global metrics hinted at good farm-level performance when feeding the wake model with the SCADA-derived inflow. The binned metrics, on the other hand, tell a less favourable story. The farm mean error generally oscillates between roughly $-5\,\%$ and $+5\,\%$ across most bins, partly averaging out over the wind rose. At around $\approx 10\,\%$ on average, the farm-MAE is more in line with the RMSE values reported in Figs. 10 and 11. This agreement nevertheless degrades sharply for some conditions, reaching more than $20\,\%$ at Anholt under very stable conditions or highly heterogeneous inflow. While this confirms the ability of engineering wake models to capture general power production trends, it also highlights their limited fidelity when finer-grained conditions are considered. This limitation should be kept in mind given the growing interest in using such models for applications that demand turbine-

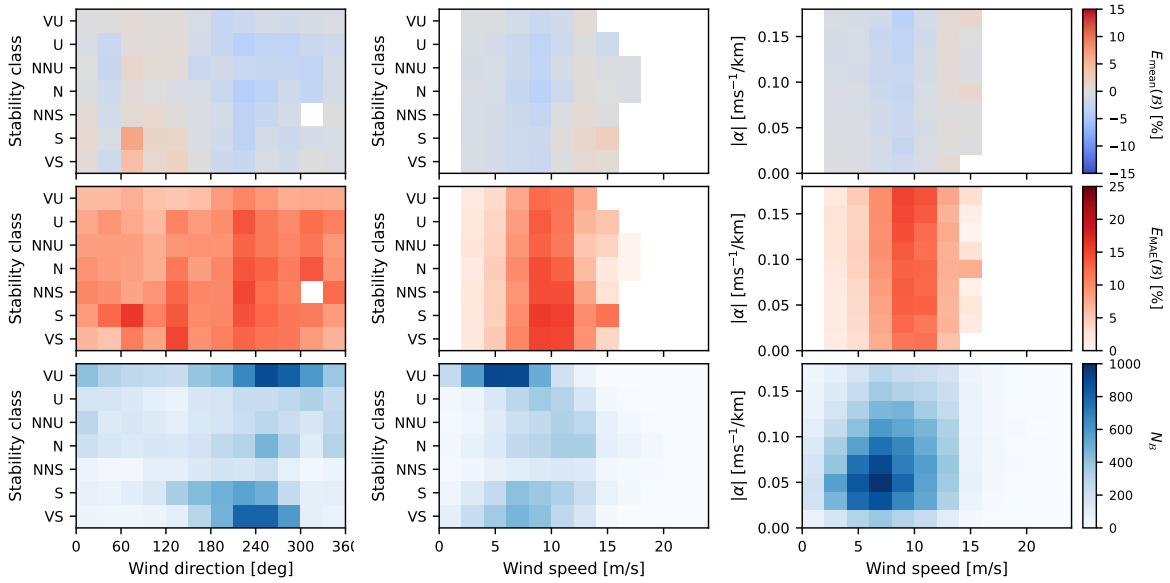


Figure 12. Westermost Rough, binned power error: farm mean E_{mean} **(Top row)** and farm-MAE norm E_{MAE} **(Middle row)** [%], and sample count $N_{\mathcal{B}_i}$ **(Bottom row)**, as a function of **(Left)** wind direction vs. stability class, **(Center)** wind speed vs. stability class, and **(Right)** wind speed vs. gradient magnitude $|\alpha|$.

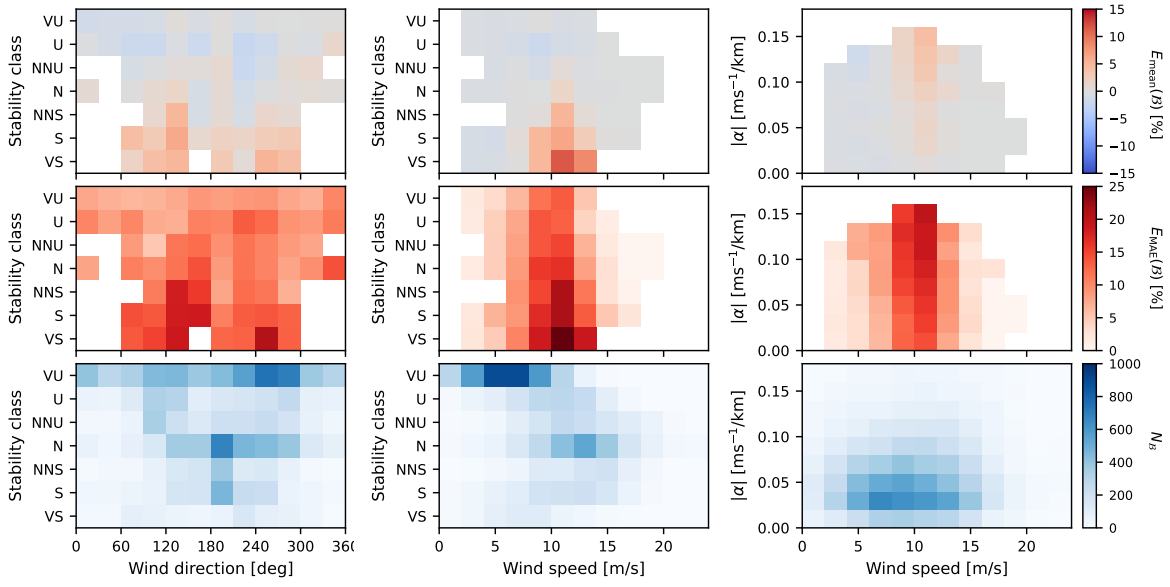


Figure 13. Anholt, binned power error: farm mean E_{mean} **(Top row)** and farm-MAE E_{MAE} **(Middle row)** [%], and sample count $N_{\mathcal{B}_i}$ **(Bottom row)**, as a function of **(Left)** wind direction vs. stability class, **(Center)** wind speed vs. stability class, and **(Right)** wind speed vs. gradient magnitude $|\alpha|$.

level or temporally-resolved power estimates, such as adequacy studies (*e.g.* O'Malley et al. 2024) or fatigue load assessment (*e.g.* Frandsen 2007; Patel et al. 2026).

Both farm-mean and farm-MAE errors concentrate in the [8-12] $\mathrm{m\,s^{-1}}$ wind speed range, which corresponds to the wake-loss regime near rated operation, where small velocity errors translate into the largest power errors. Below this range, power production is low, leading to high relative errors but negligible absolute errors. Above it, turbines transition to rated operation and power output becomes essentially independent of small wind speed variations. The error around rated conditions is nevertheless smaller than in previous investigations by the authors (Lejeune et al., 2026), which relied on publicly available power curves. This improvement is largely due to the use of SCADA-derived power curves which are intrinsically more consistent with the rest of the workflow. This stresses again the importance of methodological coherence across the full processing chain.

5.3.1 Impact of stability

For both wind farms, Figures 12 and 13 reveal a dependence of the wake model error on atmospheric stability. A similar trend is visible at both wind farms with power bias increasing when moving from unstable to stable conditions. Unstable stratification enhances turbulent entrainment and therefore accelerates wake recovery, leading to higher actual power than predicted (Hansen et al., 2012; Platis et al., 2022). Conversely, stable stratification suppresses entrainment and slows wake recovery, leading to lower actual power than predicted. Given the clear dependence of the WRF-derived TI on stability, the wake models can be considered stability-dependent. The results however show, that the observed dependency of losses to the atmospheric stability exceeds its parameterization in the models. This might be addressable with a site-specific calibration of the wake model recovery rate parameters or the access to more accurate TI estimates.

Some key differences appear between the two sites. First, the stability dependence is more pronounced at Anholt, reaching up to $E_{\mathrm{mean}} \approx 15\%$ under very stable conditions, which we partially attribute to its larger number of turbines and the resulting stronger cumulative wake effects. A second notable difference emerges when binning the data by wind direction and stability. At Anholt, the impact of stability is approximately uniform across wind direction and wind speed bins. At Westermost Rough, the picture is more contrasted. For the offshore sector [60-180°], the error shifts from positive (overestimation in VS) to near zero (in VU), while for the land sector [210-330°] it shifts from slight underestimation (in VS) to stronger underestimation (in VU). The relative impact of stability is therefore preserved across sectors, but the absolute error level differs, hinting at a direction-dependent offset. Likewise, while the farm-MAE is uniform across stability bins at Westermost Rough, it shows a strong stability dependence at Anholt reaching up to 25% in very stable conditions for the [10-12 $\mathrm{m\,s^{-1}}$] bin. This is again consistent with the higher number of turbines and the stronger sensitivity of cumulative wakes to a misrepresented recovery rate.

5.3.2 Impact of flow heterogeneity

The first observation when binning by wind speed and gradient magnitude is the difference in the total count distribution between the two farms (bottom-right panels of Figs. 12 and 13). At Anholt, the distribution is more compact, consistent with its larger spatial extent and longer distance from shore. At Westermost Rough, larger gradient magnitudes are reached, partly

because fewer WRF grid points are included in the smaller farm footprint and partly because of the closer proximity to the coast.

The farm mean error E_{mean} is mostly insensitive to gradient magnitude at Westermost Rough and shows some dependence to it at Anholt. The farm-MAE panel, on the other hand, shows a clear transition toward higher E_{MAE} with increasing gradient at Anholt, from approximately $10\,\%$ at low $|\alpha|$ to $20\,\%$ at high $|\alpha|$ in the rated region. A similar but weaker trend is observed at Westermost Rough, from approximately $7.5\,\%$ to $12.5\,\%$. The weaker response at Westermost Rough, despite its larger absolute gradients, largely reflects the difference in farm size: along its main axis, Anholt extends over more than three times the distance of Westermost Rough, so a given gradient produces a larger wind speed difference across the farm. Even at zero gradient, the farm-MAE remains high, indicating that inflow heterogeneity is not the dominant driver of the wake model error, though it does contribute to it.

The need to account for inflow heterogeneity is confirmed by Figs. 14 and 15 which report the spatial distribution of the per-turbine time-averaged power error, $e_{\mathrm{wt}} = \overline{\Delta P_{\mathrm{wt}}}$, for specific wind direction bins and compares it to the per-turbine time-averaged wind speedups, $\mu_{\mathrm{wt}} = \overline{u_{\mathrm{wt}}/u_c}$, as extracted from the WRF runs. The wind speedups are obtained by normalizing the wind speed retrieved at the turbine location u_{wt} by the wind speed at the barycenter of the wind farm, u_c. The analysis focuses on the [8-12 $\mathrm{m\,s^{-1}}$] wind speed range, where gradients have the strongest impact on the power error. For each farm, results are shown for four wind direction bins: [30-60°], [120-150°], [210-240°], and [300-330°]. Each scatter is complemented by a directional indicator reporting, for each bin, the extent of the bin studied, the mean wind direction of the inflow $\bar{u}$, the mean wind-speedups gradient direction, $\nabla\mu_{\mathrm{wt}}$, and the mean power error gradient direction, ∇e_{wt}. Consistent with the wind speed gradient definition, the power error and wind-speedups gradients are obtained by projecting the per-turbine quantity onto the WRF grid and evaluating the mean spatial gradient over the farm footprint.

Two distinct idealized power-error patterns can be distinguished from these maps as reported on Fig.16. The first appears when the power-error gradient and the wind-speedups gradient are aligned but of opposite sign (Fig.16 (a)). This configuration emerges when the wake model fails to capture the inflow heterogeneity: the wind-speedups gradient points toward the region of increasing wind speed, where the homogeneous inflow underestimates the actual wind speed, leading to an underestimation of the power on that side of the farm and vice versa. The second pattern (Fig.16 (b)) arises when the power-error gradient is aligned with the wind direction with negligible wind speed gradient, which is a signature of wake superposition effects. In this case, the error is dominated by the downstream accumulation of wake losses that are not properly captured by the wake model, producing an along-wind error gradient. In practice, the two effects are often superimposed and modulated by their respective strengths (Fig.16 (c)), leading to more complex error patterns as already noted by *e.g.* Ahsbahs et al. (2018)

Several general comments can be made based on Figs. 14 and 15. First of all, at both farms and across all sectors, the lowest per-turbine errors are consistently found on the upstream row. This is a direct consequence of the SCADA-based inflow derivation procedure: the freestream wind speed is estimated from the upstream turbines, so that once fed back through the SCADA-derived power curve, these turbines yield near-zero bias by construction. The error then grows moving downstream, reflecting the progressive accumulation of wake model inaccuracies. This is not the case for the WRF-driven results (not

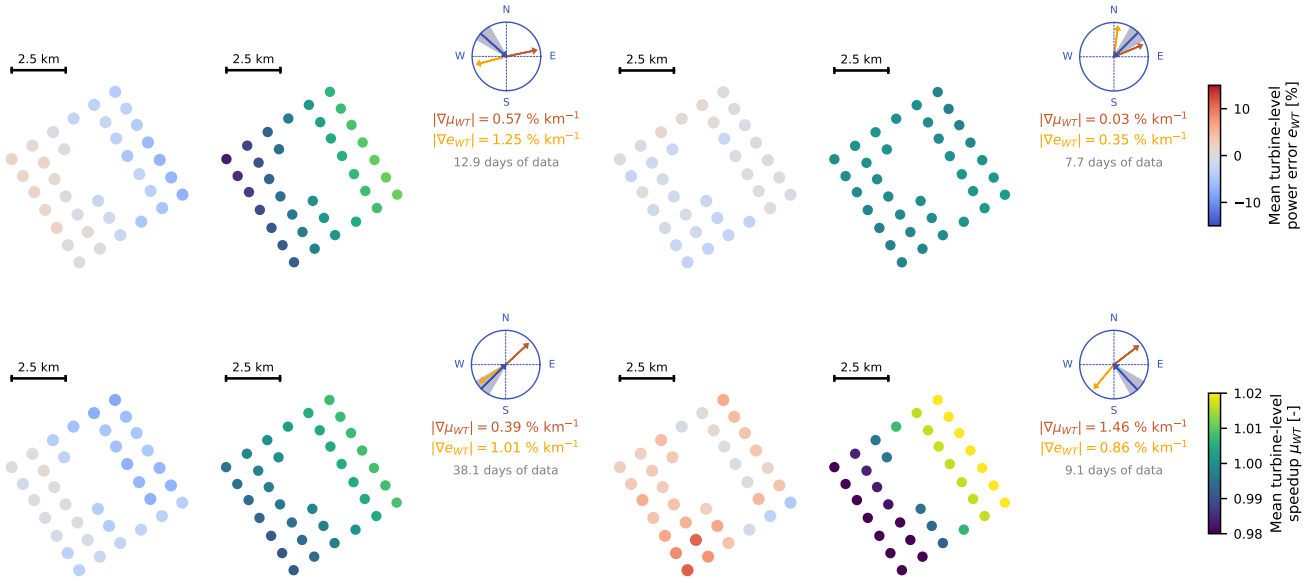


Figure 14. Westermost Rough, per-turbine mean power error, e_{wt}, for $u \in [8, 12]\,\mathrm{m\,s^{-1}}$, model-averaged, for four wind direction sectors : 30-60°, 120-150°, 210-240°, and 300-330°. Each panel includes a directional indicator showing the sector extent (shaded blue area), the mean wind direction (blue arrow), the mean wind speedups gradient direction (red arrow), and the mean power error gradient direction (yellow arrow), together with the corresponding mean wind speedups gradient and power-error gradient magnitudes.

shown), where the error is more uniformly distributed across the farm with no clear upstream-downstream pattern and an overall degraded performance since the WRF inflow is extracted at the barycenter of the wind farm.

Additionally, a remarkable agreement is observed between the WRF-derived wind-speedups and the SCADA-derived power error across most sectors and for both wind farms. This confirms that, despite being only a numerical proxy for the real data, WRF still captures the large-scale spatial trends relevant to wind farm performance. These observations however slightly vary depending on the farm considered.

Westermost Rough (Fig. 14) experiences strong coastal gradients due to its proximity to the coast with limited wake effects due to it relatively small size. Only the offshore sector [30-60°] is associated with homogeneous wind speeds: it yields a negligible wind speed gradient ($|\nabla\mu_{\mathrm{wt}}| = 0.03\,\%\,\mathrm{km^{-1}}$), yet a clear along-wind power error gradient persists ($|\nabla e_{\mathrm{wt}}| = 0.35\,\%\,\mathrm{km^{-1}}$). This gradient shows near-zero bias for upstream turbines but shifts toward negative values downstream, likely resulting from improper wake superposition as described by the simplified case (b). As visible from the speedup maps, the three remaining sectors all experience a strong coastal gradient ($|\nabla\mu_{\mathrm{wt}}| \geq 0.39\,\%\,\mathrm{km^{-1}}$), resulting in a high power error gradient over the wind farm. The [120-150°] and [300-330°] sectors correspond to the simplified case (a) identified earlier with wind-speedups and power error gradients of opposite sign. Their power error then appears clearly driven by the lack of heterogeneous inflow in the proposed wake modeling setup. The onshore sector, [210-240°], on the other hand, presents a similar

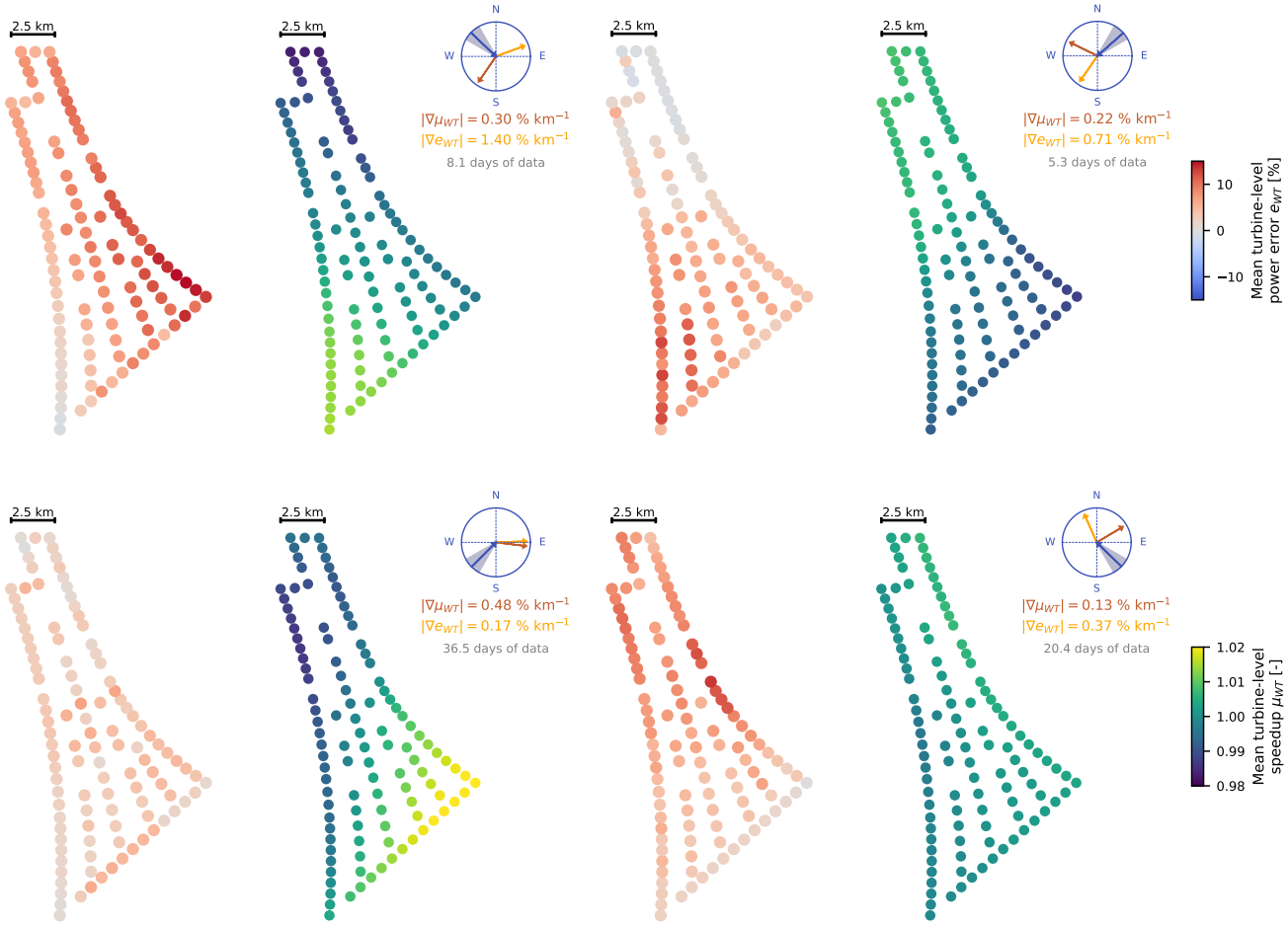


Figure 15. Anholt, per-turbine mean power error, e_{wt}, for $u \in [8, 12]\,\mathrm{m\,s^{-1}}$, model-averaged, for four wind direction sectors : 30-60°, 120-150°, 210-240°, and 300-330°. Each panel includes a directional indicator showing the sector extent (shaded blue area), the mean wind direction (blue arrow), the mean wind speedups gradient direction (red arrow), and the mean power error gradient direction (yellow arrow), together with the corresponding mean wind speedups gradient and power-error gradient magnitudes.

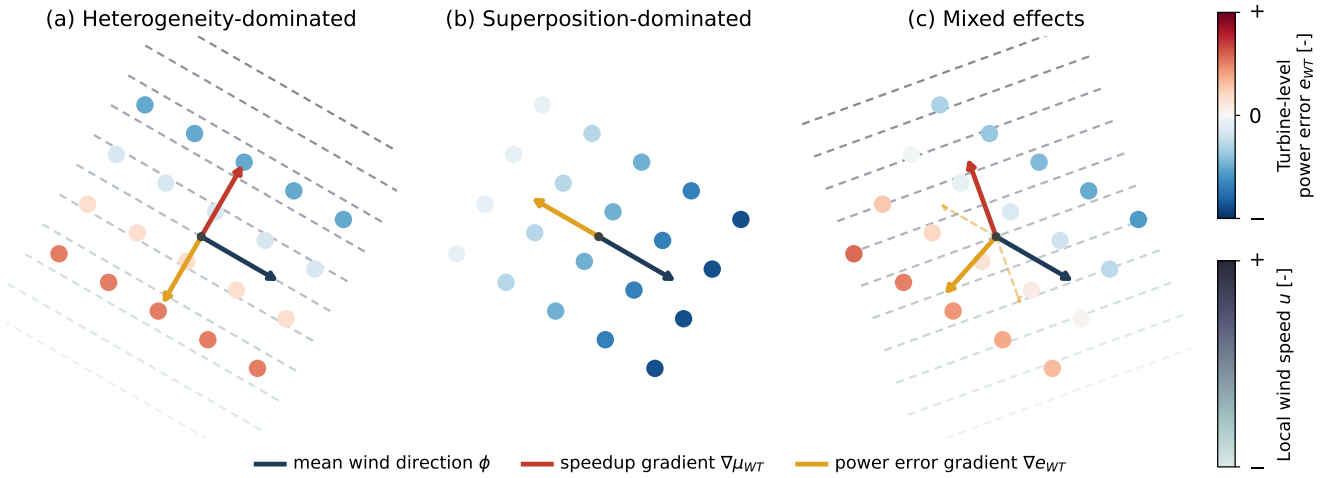


Figure 16. Idealised power-error patterns for a theoretical wind farm. Markers denote turbines coloured by their mean power error e_{wt}, positive values indicating power overestimation. Dashed lines are iso-speed contours of the background wind speed field, coloured by the local wind speed u. The mean wind direction ϕ, the wind speedups gradient $\nabla\mu_{\mathrm{wt}}$, and the resulting power error gradient ∇e_{wt} are shown as full arrows. **(a)** ∇e_{wt} aligned with $\nabla\mu_{\mathrm{wt}}$ but of opposite sign, the signature of an inflow heterogeneity dominated error. **(b)** Uniform inflow, with ∇e_{wt} aligned with the wind axis as wake losses accumulate downstream. **(c)** Both mechanisms acting together.

configuration but aligned with the main wind direction, thus potentially adding a wake modeling bias on top of the one induced by flow heterogeneity, corresponding to (c).

Figure 17 extends this comparison from the four selected bins to the full wind rose. It demonstrates quantitatively the link between the derived wind speedup gradient and the power error gradient across all wind direction bins. Both the along- and cross-wind components of the wind speedup and of the power error show a similar evolution across the wind rose, confirming that flow heterogeneity acts on top of wake modeling as one of the main driver of the spatial distribution of the power error.

At Anholt (Fig. 15), a strong spatial correlation is again observed between the speedups and power error maps. This correlation however does not translate quantitatively into the metrics studied. All sectors demonstrate some degree of heterogeneity, with wind-speedups gradients spanning 0.13 to $0.48\,\%\,\mathrm{km}^{-1}$ against 0.03 to $1.46\,\%\,\mathrm{km}^{-1}$ at Westermost Rough. For all sectors but the [300-330°] one, the power error gradient is roughly aligned with the wind direction but of opposite sign hinting that some of the power losses can be blamed on the intrinsic limitation of the wake model in computing deep-array effects. This is particularly apparent for the south-westerly sector [210-240°], which combines the largest wind speed gradient of the farm ($|\nabla\mu_{\mathrm{wt}}| = 0.48\,\%\,\mathrm{km}^{-1}$) with the smallest power error gradient ($|\nabla e_{\mathrm{wt}}| = 0.17\,\%\,\mathrm{km}^{-1}$), suggesting that the heterogeneity and wake superposition effects partially compensate each other for this sector. Conversely, for the north-westerly sector [300-330°], a marked power overestimation is observed on the south-west side of the farm transitioning to underestimation on the north-east side, consistent with the direction of the wind speed gradient and similarly to what was observed at Westermost

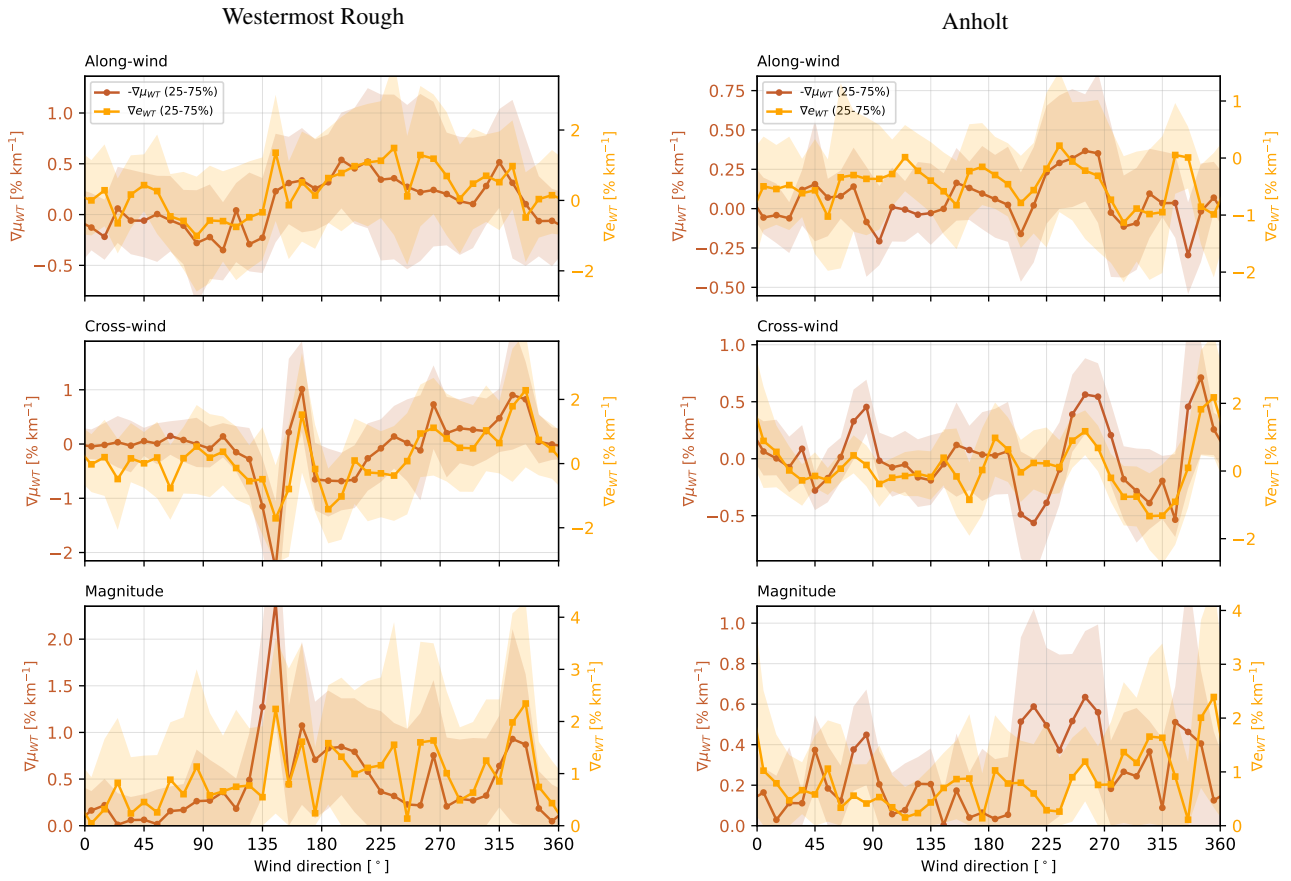


Figure 17. Directional distributions of the wind-speedup and power-error gradients, for $u \in [8, 12]\,\mathrm{m\,s^{-1}}$, model-averaged: $-\nabla\mu_{\mathrm{wt}}$ (dark red, left axis) and ∇e_{wt} (orange, right axis), both in $\%\,\mathrm{km^{-1}}$. Median (full line) and P25 and P75 (shaded area). **(Top)** along-wind component. **(Center)** cross-wind component. **(Bottom)** gradient magnitude.

Rough. These observations are consistent with past investigations led at Anholt, with Peña et al. (2018) reporting that explicitly accounting for the horizontal wind-speed gradient in AEP calculations had a limited impact on the total yield at Anholt, though affecting the individual turbine power error.

Figure 17 is less conclusive for Anholt than it was for Westermost Rough. The correlation between the two binned sequences largely erodes despite the good visual match between the speedups and power error maps observed on Fig. 15. The link between the wind-speedups gradient and the power error gradient can nonetheless be spotted for the cross wind component over the [225-60°] window with coinciding peaks over the two sequences. Several causes explaining this degraded match could be hypothesized: a less Cartesian layout, more intricate wake interactions, or the more complex structure of the flow caused by the larger spatial extent of the wind farm. All of these ultimately challenge the linear nature of the proposed approach, making

it less adequate for this specific case rather than truly questioning the actual impact of the heterogeneity on the wake model accuracy.

Taken together, the spatial maps confirm that wake superposition and inflow heterogeneity are two distinct mechanisms contributing to the per-turbine error: the former produces a downstream error pattern aligned with the wind direction, while the latter produces a flow asymmetry tied to the coastal gradient. The wake superposition is a long-standing limitation of analytical wake models, already studied by many authors (Zong and Porté-Agel, 2020; Lanzilao and Meyers, 2021). On the other hand, capturing inflow heterogeneity requires either a spatially-resolved inflow source such as WRF (Nygaard et al., 2022), a flow model that resolves the coastline explicitly van der Laan et al. (2017) or space- and time-resolved inflow maps from dedicated field campaigns (Ahsbahs et al., 2018).

The coexistence of both mechanisms has direct implications for calibration. As illustrated at Anholt, heterogeneity and wake superposition errors can partially compensate each other. A wake expansion coefficient k_{wk} tuned on such a sector would mix together two opposing biases and is unlikely to transfer to other conditions. The south-easterly sector of Westermost Rough [120-150°] raises another concern. There, the observed gradient is caused by both the land-sea roughness transition and the wake of the neighbouring Humber Gateway wind farm. A single gradient indicator cannot split these contributions. These observations motivate the need for heterogeneity-aware calibration strategies that distinguish coastal effects from farm-to-farm interactions.

5.4 Temporal distribution of power error

Although AEP has historically been the primary performance metric for wind farm assessment, the industry is increasingly shifting toward time-series-based approaches. This shift reflects evolving operational concerns, notably the growing relevance of grid integration and balancing requirements (*e.g.* O'Malley et al. 2024). Figure 18 examines the temporal accuracy of the wake model predictions by binning the farm mean power error per 2 h blocks and per month. For each farm, three panels are reported: the power error obtained with the SCADA inflow, the same quantity for the WRF-derived inflow, and the mean stability class according to WRF. It should be noted that the available data spans two years for Westermost Rough but less than one year for Anholt so that the statistics presented below may not be fully representative of a typical meteorological year. The goal of this section is not to exhaustively discuss the climatology of these sites but rather to quantify what their seasonal and daily variability implies for the accuracy of the wake modeling workflows.

Since the conditions encountered are not uniform across the year, the error does not average out when binned on a temporal basis, resulting in varying power biases fluctuating with the atmospheric conditions. At Westermost Rough, the summer months are associated with stable conditions, the sea surface remaining colder than the above air. This triggers less wake recovery and hence stronger wake losses, a mechanism not fully captured by the wake models as already evidenced in Sec. 5.3.1. A similar but weaker pattern emerges on the diurnal axis: in summer, night-time is associated with lower air temperatures while the sea temperature remains mostly constant triggering close to neutral conditions. The resulting variation in wake losses is fully captured by neither the SCADA nor the WRF-driven setup with respective time-binned errors spanning $-5\,\%$ to $+2\,\%$ and $-7.5\,\%$ to $+10\,\%$. These values are significantly higher than what would be expected from an AEP perspective and may

seriously hinder the ability of wake model to deal with applications where temporal accuracy is required. The model behaves best in the mid-season where the bias is negligible and worst in winter when the relatively warm sea triggers unstable conditions. The Anholt site is characterized by a different wind climatology, yet the same general conclusions hold regarding the behavior of the wake model though more extreme bias fluctuations are encountered over the year: -2.5 to $+3\,\%$ and -10 to $+3\,\%$

On top of being higher than its SCADA-counterpart, the WRF-derived power bias is especially interesting to investigate. Indeed, the same WRF-derived turbulence intensity is used to generate both the WRF- and SCADA-derived inflows, so that the difference between the two upper panels originates solely from the estimations of the wind speed and direction of the inflow. Additionally, the binning is purely temporal and does not involve stability in any way. Yet, a strong visual correlation persists between the patterns of the middle and bottom panels indicative of a stability-dependent accuracy of the WRF inflow.

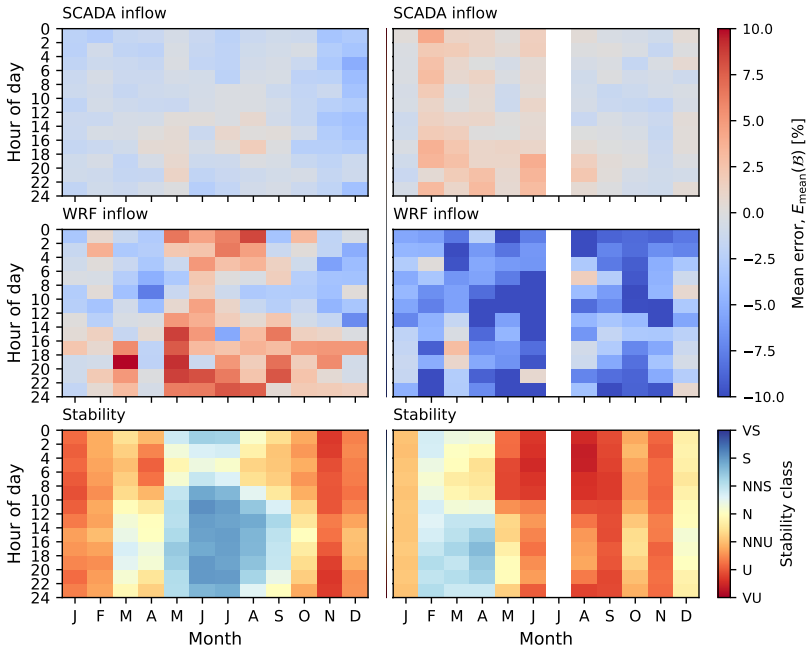


Figure 18. Temporal distribution of power error: **(Left)** Westermost Rough. **(Right)** Anholt. **(Top)** farm mean power error for the SCADA-derived inflow [%]; **(Middle)** farm mean power error for the WRF-derived inflow [%]; **(Bottom)** corresponding atmospheric stability class as retrieved from WRF. Both power error histograms are shown as a function of month and 2 h period of the day and model- and software-averaged.

6 Discussion

Connecting back to the initial identification of uncertainty sources in the modelling workflow, the results of Sects. 4 and 5 allow their relative importance and degree of interaction to be assessed. A visual synthesis is additionally provided by Fig. 19,

which reports the distribution of the per-timestep farm power bias for every combination of inflow source, turbulence intensity treatment, wake model and software implementation. Complementary to the simulations driven by the synthetic WRF turbulence intensity, results obtained with a fixed value of 6 % are also reported. Both the timeseries-based median and mean are reported on this figure, the former being equal to the values reported in the power performance matrices.

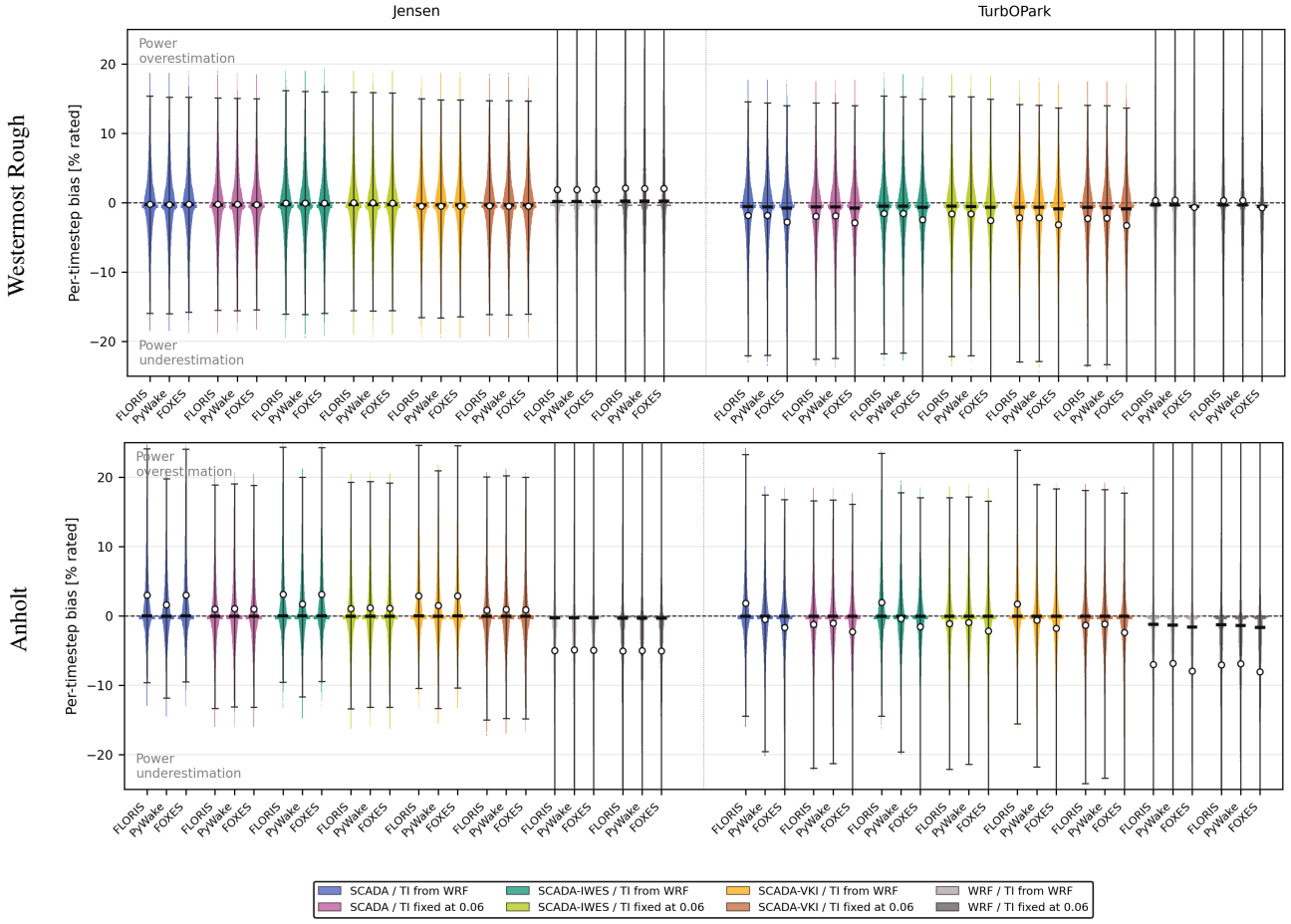


Figure 19. Binned bias distribution for different combination of inflow/wake model/software package: violin plot of the timeseries of the power biases. Median (thick black line) and mean (white circle) of the timeseries bias. Upper and lower horizontal mark indicate the 5% and 95% bounds. **(Top)** Westermost Rough and **(Bottom)** Anholt.

6.1 Inflow uncertainties

The two SCADA-filtering workflows produce nearly interchangeable wind speed timeseries ($b \leq 0.04\,\mathrm{m\,s^{-1}}$, $r \approx 1$), confirming that this source of uncertainty is small for wind speed. Wind direction, by contrast, shows larger discrepancies at Anholt ($b = 7.27°$), most likely attributable to differences in northing calibration. Accordingly, Fig. 19 shows a good agreement between the power bias distributions obtained from the SCADA-derived inflows. The consistency of the wind speeds fed to the turbine power curves explains this match, wind direction acting only as a secondary effect through the wake propagation.

Comparable levels of wind direction bias are found between WRF and the SCADA-derived inflows, whereas the WRF-derived wind speed diverges more markedly from SCADA, with biases up to $|b| = 0.50\,\mathrm{m\,s^{-1}}$, a lower correlation and a higher RMSE indicative of a temporal mismatch. As power scales with the cube of the wind speed, minor wind speed errors turn into large power deviations, which directly translates into the distribution of the power error. While the bias remains close to zero on average, the uncertainty envelope associated with the WRF-derived inflow is much larger than for the SCADA one.

Finally, the choice of TI treatment appears to be a second-order effect, not significantly affecting the time-based distribution of the power bias. Still, throughout the investigations and compared to past studies led by the authors (Lejeune et al., 2026), the inclusion of TI was observed to have a small but beneficial impact on the wake model estimates, especially when binning the data based on stability.

6.2 Wake-model uncertainties

Neither model (*i.e.* Jensen and TurbOPark) was consistently shown to outperform the other across all conditions. Both are ultimately limited by their underlying assumptions, in particular by expansion coefficients that lump together a wide range of physical mechanisms, from stability-driven recovery to flow heterogeneity. The choice of wake model nevertheless stands as a major contributor to the overall bias, with differences of up to $5\,\%$ observed at Anholt between the two models for the same inflow source and software package. It is however not clear to which extent these differences could be reduced through an adequate calibration of the underlying models.

6.3 Implementation uncertainties

Jensen exhibits consistent performance across FLORIS, PyWake and FOXES, confirming the robustness of this simple model. TurbOPark, by contrast, shows inter-software bias spreads of up to $4\,\%$ at Anholt for identical inputs. This fundamentally questions the transferability of calibration efforts across software implementations.

6.4 Condition-dependent uncertainties

The binned analysis reveals that the time-averaged bias alone is not sufficient to assess wake model performance. On Fig. 19, the mean and, perhaps even more remarkably, the median both stand close to zero for most configurations, yet they reflect widely different realities depending on the wake model considered, as shown by the varying spread of the error across wake modeling workflows. Farm-mean errors of $\pm 5\,\%$ and farm-MAE exceeding $20\,\%$ were indeed observed under specific atmo-

spheric conditions, despite near-zero wind farm biases. Atmospheric stability is the dominant driver, with a clear monotonic trend from negative error under unstable conditions to positive error under stable conditions, while flow heterogeneity acts as a secondary effect. The temporal analysis confirms that these condition-dependent biases translate into systematic seasonal and diurnal patterns in the power error.

Overall, the results show that uncertainties related to operating conditions and inflow dominate the overall error, whereas the selected SCADA filtering workflow has only a limited effect on the estimated power output. This finding should, however, be interpreted with caution, since the close agreement between the filtering workflows partly results from systematically excluding problematic timestamps when calculating the error metrics, as described in Sect. 5.1.

Two additional results stand out. First, the choice of software package can be as impactful as the choice of wake model itself: an uncertainty most often invisible to the modeller. Second, bias alone is not sufficient to characterise model performance: a bias-based analysis would incorrectly conclude that the WRF-derived inflow provides the most faithful inflow at Westermost Rough on an AEP basis, whereas the corresponding binned-bias distributions (Fig. 19) are markedly broader than their SCADA-driven counterparts. For these specific cases, the near-zero mean results occurs thanks to compensating errors across atmospheric conditions rather than from genuine accuracy.

7 Conclusions

This work set out to address a question that arises when a wake model is compared to measurements: *when a model disagrees with observations, how much of the discrepancy originates from the model physics, how much from the user's implementation choices, and how much from the inflow and observational data used as inputs?* By combining four inflow derivation procedures (SCADA-IWES, SCADA-VKI, WRF and LiDAR), two wake models sensitized to ambient TI (Jensen and TurbOPark) and three software packages (PyWake, FOXES or FLORIS) across two offshore wind farms (Westermost Rough and Anholt), the present study attempts to quantitatively assess these contributions.

Inflow conditions were found to have the strongest influence on power estimates, particularly at individual time stamps and for binned data. The two SCADA-filtering workflows produce nearly interchangeable wind-speed timeseries, while wind direction exhibits slightly larger differences attributable to the nacelle-alignment methodology used. Harmonising the two filtering workflows proved difficult in the absence of community guidelines, particularly concerning the definition and use of the reference operational flags. This highlights another key requirement wake modeling workflow should enforce: the consistency of the workflow as a whole. For instance, improved results were obtained when the power curves and driving inflow were derived from the same dataset. Likewise, the location at which inflow velocities are extracted must remain consistent with the modeling assumptions as it might vary depending on the inflow.

Despite being a proxy for the real flow, WRF demonstrated that it reproduces the large SCADA trends remarkably well. This was especially well illustrated by the correspondence between the WRF speedups and the measured power error. However, the temporal correlation of the mesoscale data proved to be too low to derive robust power estimates for individual timestamps

or even binned-data. Yet, for some configurations, the WRF-derived inflow appeared to be the most accurate wake modeling workflow from an AEP (*i.e.* time-averaged bias) perspective while RMSE and Pearson corelation immediately discarded this initial conclusion. This strongly suggests moving away from AEP as the sole evaluation metric as compensation of biases across conditions can hide problems in representativeness of the inflow.

The inter-model comparison could not establish which of Jensen or TurbOPark led to more accurate results. Their accuracy actually depends on the inflow used and the farm studied and could certainly be further improved by an adequate tuning of their model parameters. The presented analysis should additionally be completed by the analysis of wake losses accuracy, effectively decoupling inflow accuracy from wake modeling accuracy. Nevertheless, TurbOPark proved to be slightly more sensitive to the software package used than Jensen, with a spread in AEP of up to 4% for identical inputs in the most extreme case. This source of uncertainty, invisible to the modeller, directly motivates ongoing community standardisation efforts such as WindIO and the WIFA framework.

Regarding the wake models, the comparison with SCADA power data under normal operation shows that their performance is not uniform across atmospheric states. Most of the error occurs in the [8-12 $\mathrm{m\,s^{-1}}$] range, where wake effects are the strongest. Directional trends are more pronounced at Westermost Rough than at Anholt, consistent with its greater distance from shore. Although both wake models feature a stability-dependent expansion coefficient, they fail at fully capturing the effect of atmospheric stability with unstable conditions leading to power underestimation and vice versa.

Flow heterogeneity was not identified as the dominant contributor to the global farm error, yet clear spatial patterns associated with it emerge when the error is evaluated at the turbine-level. Disentangling flow heterogeneity from wake recovery is thus a prerequisite for more accurate predictions. This also carries a cautionary note for wake-model tuning: calibrating the wake-expansion coefficient at the bin level may still be insufficient to reproduce the actual farm behaviour because the calibration may incorrectly attribute an inflow gradient to wake recovery. These calibration efforts are further undermined by the lack of long-term SCADA datasets. When binning the data per atmospheric conditions, the user quickly experiences data scarcity, preventing any statistically meaningful conclusion from being drawn. Isolating any single effect therefore remains challenging as two years of concurrent SCADA data are insufficient to populate all relevant combinations of atmospheric states.

Together, these findings call for improved community guidelines and shared validation datasets to support more robust and reproducible wake-modelling workflows. They additionally support a shift from classical AEP-based validation toward timeseries- or binned-based approaches. This should then allow to better isolate the phenomena driving the power production error at the turbine-level thereby opening the door to more successful calibration efforts.

Further work will build on these results by developing a calibration framework toward site-specific approaches, in particular through the explicit inclusion of neighboring wind farm wakes, coastal gradients and atmospheric stability effects.

Appendix A: Wake model configuration

The windIO configuration used to drive the wake simulations is provided in Tab.A1. These parameters are processed and interpreted by WIFA that eventually feeds them to the underlying wake modeling software packages (PyWake, FOXES or

FLORIS) along with the inflow timeseries. How these parameters and states are interpreted and processed by WIFA or its underlying tools however falls outside the scope of the present paper.

Table A1. Wake model configuration passed to FLORIS, PyWake and FOXES through the WindIO interface. Keys refer to the WindIO wind farm model schema. Settings are identical between the two models except where indicated.

Component	WindIO key	Jensen	TurbOPark
Wind deficit	`wind_deficit_model.name`	Jensen	TurbOPark
	`...k_a`	0.00	0.00
	`...k_b`	0.67	0.04
	`...free_stream_ti`	`true`	`false`
	`use_effective_ws`	`false`	
Axial induction	`axial_induction_model`	1D	
Deflection	`deflection_model.name`	None	
Turbulence	`turbulence_model.name`	CrespoHernandez	
Superposition	`...ws_superposition`	Squared	
	`...ti_superposition`	Squared	
Rotor averaging	`rotor_averaging.grid`	grid	
	`n_x/n_y_grid_points`	3×3	
	`background/wake_averaging`	grid	
	`.._exponent_for_power`	1	
	`.._exponent_for_ct`	1	
Blockage	`blockage_model.name`	None	

Appendix B: Turbulence intensity cross-comparison

None of the sources considered provides a reliable estimate of the ambient TI. Nacelle-mounted anemometers sit downstream of the rotor and are therefore affected by rotor-induced turbulence and flow distortion perturbing their TI measurements. Following IEC 61400, these measurements should accordingly be corrected through a nacelle transfer function, unavailable here, so that the uncorrected SCADA-derived TI are expected to overestimate the ambient TI compared to freestream references. The vertical profiling LiDAR also suffer from limitations for TI measurements: on top of being partially waked, they systematically misestimate turbulence owing to probe-volume averaging and the cross-contamination effect induced by the conical scanning geometry, as demonstrated by Kelberlau and Mann (2020). Finally, with no explicit sub grid-scale turbulence modelling, the WRF-derived TI cannot be considered as the baseline reference either.

These observations are confirmed by the correlation plots and performance matrices of Fig. B1. The inter-SCADA comparison shows nearly no bias and excellent correlation ($r \approx 0.95$) at both sites. When compared to the LiDAR, the SCADA-derived TI shows a negative bias of -3 % to -9 %, as a result of the local flow features affecting the nacelle anemometry. The WRF-

derived TI shows a consistent negative bias of -2 % to -5 %, except for the unwaked Anholt case where the bias reaches +2 %. On the LiDAR-WRF correlation scatter, the waked measurements provide higher TI estimates than the freestream WRF, which is not affected by wake-added TI; once waked conditions are excluded, the scatter shifts back toward the 1:1 line, suggesting that the WRF-derived TI is a reasonable estimate of the ambient TI. Correlation and RMSE nevertheless remain poor across pairwise comparisons, reflecting the difficulty of using a uniform definition of TI across sources and the sensitivity of TI to local flow features.

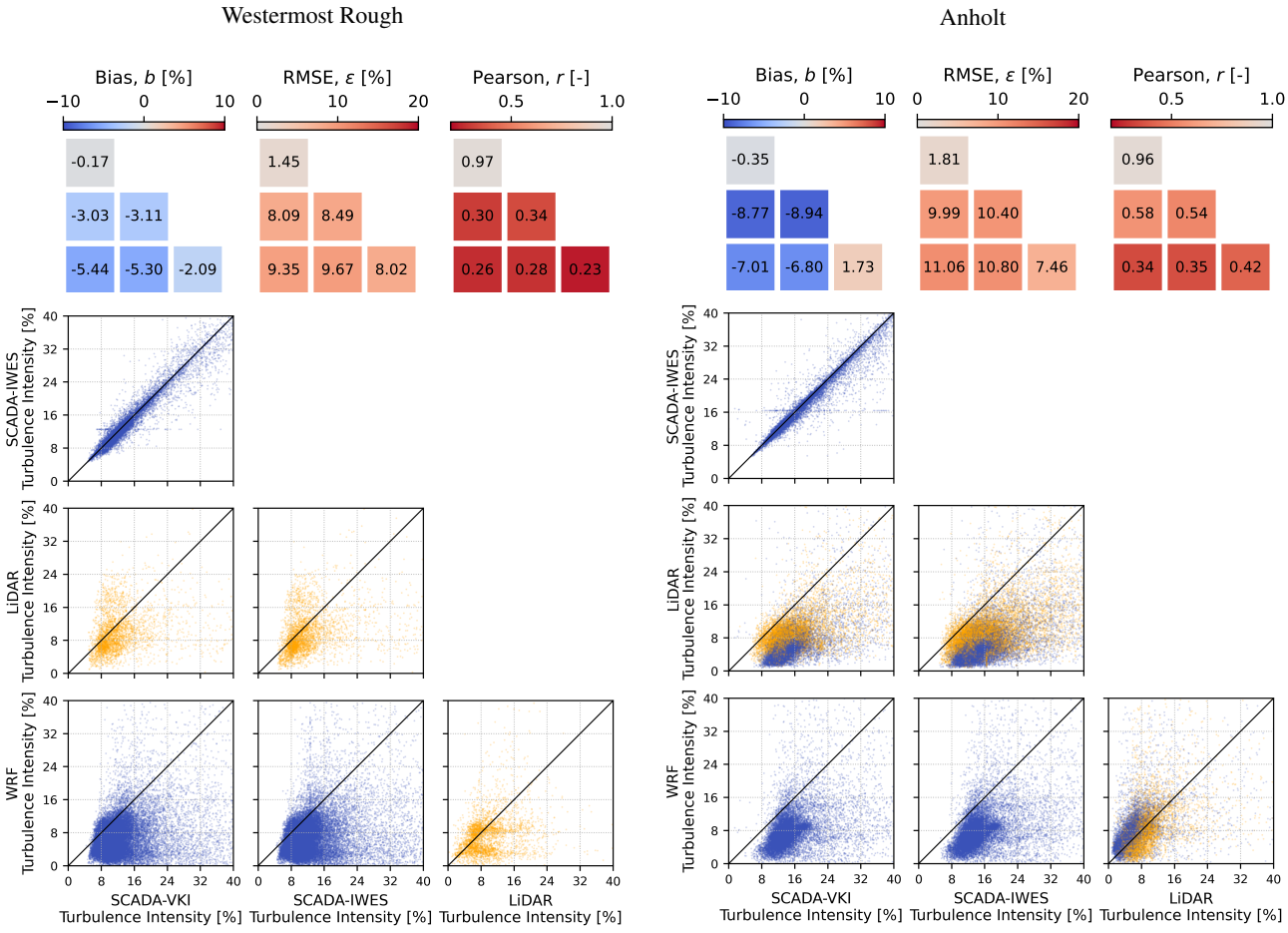


Figure B1. Turbulence intensity performance matrices and correlation plots for **(Left)** Westermost Rough and **(Right)** Anholt. The cells of the performance matrices follow the same ordering as the correlation plots. **(Top per panel)**. Pairwise performance matrices: b [%], ε [%], and r [-] across all inflow sources. Each cell shows the metric for the y-source relative to the x-source. **(Bottom per panel)** Pairwise correlation scatter plot; blue: unwaked conditions; orange: waked conditions.

Code and data availability. The SCADA data used in this study are subject to confidentiality agreements with the wind farm operators and cannot be publicly released. Authors' contributions to WindIO and WIFA have been made available through their respective open-source repositories. The DTWO project datasets released to the public are described at https://dtwo-project.eu/useful-resources/.

Author contributions. LA, LV, ML, NC, and WM conceptualized the study. LV, ML, and NC carried out the SCADA-data filtering and analysis, while ML developed the comparison framework and performed the engineering wake-model simulations. All authors contributed to the interpretation of the results. ML drafted the manuscript, and all authors reviewed and edited it.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work was carried out within the DTWO (Federated Digital Twins for Wind-Offshore) project, funded by the European Union's Horizon Europe research and innovation programme under grant agreement No. 101146689. The authors thank Ørsted for granting access to the SCADA data, and Nicolai Nygaard and Lina Poulsen from Ørsted for their interest in this work and their careful review. Implementation was partially assisted by GitHub Copilot. Generative AI was additionally used to review the final draft and to reformulate some sections ensuring that such changes did not alter the scientific content of the paper. In all cases, the authors handled the final review and validation of the manuscript.

References

Ahsbahs, T., Badger, M., Volker, P., Hansen, K. S., and Hasager, C. B.: Applications of Satellite Winds for the Offshore Wind Farm Site Anholt, Wind Energ. Sci., 3, 573–588, https://doi.org/10.5194/wes-3-573-2018, 2018.

Barthelmie, R. J. and Jensen, L. E.: Evaluation of Wind Farm Efficiency and Wind Turbine Wakes at the Nysted Offshore Wind Farm, Wind Energy, 13, 573–586, https://doi.org/10.1002/we.408, 2010.

Bastankhah, M. and Porté-Agel, F.: A New Analytical Model for Wind-Turbine Wakes, Renewable Energy, 70, 116–123, https://doi.org/10.1016/j.renene.2014.01.002, 2014.

Beltrán, J., Guerrero, J. J., Melero, J. J., and Llombart, A.: Detection of nacelle anemometer faults in a wind farm minimizing the uncertainty, Wind Energy, 16, 939–952, https://doi.org/10.1002/we.1535, 2013.

Bulder, B. H., Bot, E. T. G., and Marina, A. J.: Scoping analysis of the potential yield of the Hollandse Kust (zuid) wind farm sites and the influence on the existing wind farms in the proximity, Technical Report ECN-E–16-021, Energy Research Centre of the Netherlands (ECN), 2016.

Chen, F. and Dudhia, J.: Coupling an Advanced Land Surface–Hydrology Model with the Penn State–NCAR MM5 Modeling System. Part I: Model Implementation and Sensitivity, Monthly Weather Review, 129, 569 – 585, https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2, 2001.

Clifton, A., Clive, P., Gottschall, J., Schlipf, D., Simley, E., Simmons, L., Stein, D., Trabucchi, D., Vasiljevic, N., and Würth, I.: IEA Wind Task 32: Wind Lidar Identifying and Mitigating Barriers to the Adoption of Wind Lidar, Remote Sensing, 10, https://doi.org/10.3390/rs10030406, 2018.

commissie mer: Wind turbine power and thrust coefficients, https://commissiemer.nl, 2021.

Crespo, A. and Hernández, J.: Turbulence characteristics in wind-turbine wakes, Journal of Wind Engineering and Industrial Aerodynamics, 61, 71–85, https://doi.org/10.1016/0167-6105(95)00033-X, 1996.

Doekemeijer, B. M., Wingerden, J. V., and Fleming, P. A.: A Tutorial on the Synthesis and Validation of a Closed-Loop Wind Farm Controller Using a Steady-State Surrogate Model, in: 2019 American Control Conference (ACC), pp. 2825–2836, 2019 American Control Conference (ACC), ISSN 0743-1619, https://doi.org/10.23919/ACC.2019.8815126, 2019.

Doekemeijer, B. M., Simley, E., and Fleming, P.: Comparison of the Gaussian Wind Farm Model with Historical Data of Three Offshore Wind Farms, Energies, 15, https://doi.org/10.3390/en15061964, 2022.

DTWO Consortium: DTWO: Federated Digital Twins for Wind–Offshore, https://dtwo-project.eu/, 2024.

EMD International A/S: windPRO 4.2: ENERGY, https://help.emd.dk/knowledgebase/content/windPRO4.2/c3-UK_windPRO4.2_ENERGY.pdf, windPRO 4.2 User Manual, ENERGY module. Accessed 2026-06-15, 2025.

FGW: Technical Guidelines for Wind Turbines 10 (TG10): Determination of Site Quality after Commissioning, 2024.

Fischereit, J., Stipa, S., Du, J., Strickland, J., Lampert, A., Baerfuss, K., Schmitt, L., and Larsén, X. G.: Wind farm effects in the atmosphere and waves: a mesoscale model inter-comparison, Journal of Physics: Conference Series, 3224, 032 107, https://doi.org/10.1088/1742-6596/3224/3/032107, 2026.

Fitch, A. C., Olson, J. B., Lundquist, J. K., Dudhia, J., Gupta, A. K., Michalakes, J., and Barstad, I.: Local and Mesoscale Impacts of Wind Farms as Parameterized in a Mesoscale NWP Model, Monthly Weather Review, 140, 3017–3038, https://doi.org/10.1175/MWR-D-11-00352.1, 2012.

Fleming, P., Doekemeijer, B., and Simley, E.: FLASC (FLORIS-based Analysis for SCADA data), https://doi.org/10.11578/dc.20220511.1, 2022.

Frandsen, S. T.: Turbulence and Turbulence-Generated Structural Loading in Wind Turbine Clusters, Risoe-R 1188(EN), Forskningscenter Risoe, Denmark, publisher's PDF, version of record, 2007.

Gaumond, M., Réthoré, P.-E., Ott, S., Peña, A., Bechmann, A., and Hansen, K. S.: Evaluation of the wind direction uncertainty and its impact on wake modeling at the Horns Rev offshore wind farm, Wind Energy, 17, 1169–1178, https://doi.org/10.1002/we.1625, 2014.

Hahmann, A. N., Sīle, T., Witha, B., Davis, N. N., Dörenkämper, M., Ezber, Y., García-Bustamante, E., González-Rouco, J. F., Navarro, J., Olsen, B. T., and Söderberg, S.: The making of the New European Wind Atlas – Part 1: Model sensitivity, Geoscientific Model Development, 13, 5053–5078, https://doi.org/10.5194/gmd-13-5053-2020, 2020.

Hansen, K. S., Barthelmie, R. J., Jensen, L. E., and Sommer, A.: The Impact of Turbulence Intensity and Atmospheric Stability on Power Deficits Due to Wind Turbine Wakes at Horns Rev Wind Farm, Wind Energy, 15, 183–196, https://doi.org/10.1002/we.512, 2012.

Hong, S.-Y., Dudhia, J., and Chen, S.-H.: A Revised Approach to Ice Microphysical Processes for the Bulk Parameterization of Clouds and Precipitation, Monthly Weather Review, 132, 103 – 120, https://doi.org/10.1175/1520-0493(2004)132<0103:ARATIM>2.0.CO;2, 2004.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, Journal of Geophysical Research: Atmospheres, 113, https://doi.org/10.1029/2008JD009944, 2008.

Janjić, Z. I.: The Step-Mountain Eta Coordinate Model: Further Developments of the Convection, Viscous Sublayer, and Turbulence Closure Schemes, Monthly Weather Review, 122, 927 – 945, https://doi.org/10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2, 1994.

Jensen, N. O.: A Note on Wind Generator Interaction, Tech. rep., Risø National Laboratory, ISBN 87-550-0971-9, 1983.

Jong, R., Doekemeijer, B., Verzijlbergh, R., and Wingerden, J. W.: Validation of real-weather LES using SCADA data from a large modern offshore wind farm, Journal of Physics: Conference Series, 3224, 032 003, https://doi.org/10.1088/1742-6596/3224/3/032003, 2026.

Kain, J. S.: The Kain–Fritsch Convective Parameterization: An Update, Journal of Applied Meteorology, 43, 170 – 181, https://doi.org/10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2, 2004.

Katic, I., Højstrup, J., and Jensen, N. O.: A Simple Model for Cluster Efficiency, in: EWEC'86. Proceedings. Vol. 1, edited by Palz, W. and Sesto, E., pp. 407–410, A. Raguzzi, 1987.

Kelberlau, F. and Mann, J.: Cross-contamination effect on turbulence spectra from Doppler beam swinging wind lidar, Wind Energy Science, 5, 519–541, https://doi.org/10.5194/wes-5-519-2020, 2020.

Lanzilao, L. and Meyers, J.: A new wake-merging method for wind-farm power prediction in the presence of heterogeneous background velocity fields, Wind Energy, 25, 237–259, https://doi.org/10.1002/we.2669, 2021.

Larsén, X. G. and the DTWO Consortium: The Digital Twins for Winds-Offshore (DTWO) Project, in: EGU General Assembly 2025, pp. EGU25–13 201, Vienna, Austria, https://doi.org/10.5194/egusphere-egu25-13201, 2025.

Larsén, X. G., Imberger, M., and Floors, R.: Modelling global offshore turbulence intensity including large-scale turbulence, stability and sea state, Wind Energy Science, 11, 1853–1869, https://doi.org/10.5194/wes-11-1853-2026, 2026.

Larsén, X. G.: Calculating Turbulence Intensity from mesoscale modeled Turbulence Kinetic Energy, Technical report (dtu wind energy e-0233), DTU Wind Energy, Technical University of Denmark, Roskilde, Denmark, https://backend.orbit.dtu.dk/ws/portalfiles/portal/364980906/TKE2TI-20240627.pdf, 2024.

Lee, J. C. Y. and Fields, M. J.: An overview of wind-energy-production prediction bias, losses, and uncertainties, Wind Energy Science, 6, 311–365, https://doi.org/10.5194/wes-6-311-2021, 2021.

Lejeune, M., Vollmer, L., Coudou, N., Alcayaga, L., and Munters, W.: Toward an integrated wake model calibration tool fed with SCADA data: correlating wake model performances to the atmospheric state, Journal of Physics: Conference Series, 3224, 032 008, https://doi.org/10.1088/1742-6596/3224/3/032008, 2026.

Mayol, M., Rowen, M., and Erbslöh, S.: Validation of WRF-Derived Synthetic Turbulence Intensity for Offshore Wind Applications and Its Impact on AEP, Journal of Physics: Conference Series, 3224, 022 043, https://doi.org/10.1088/1742-6596/3224/2/022043, 2026.

Mudafort, R., Quick, J., and Schulte, J.: Comparison of Steady-State Analytical Wake Models Implemented in Wind Farm Analysis Software, Journal of Physics: Conference Series, 2767, 052 066, https://doi.org/10.1088/1742-6596/2767/5/052066, 2024.

Nakanishi, M. and Niino, H.: An Improved Mellor–Yamada Level-3 Model: Its Numerical Stability and Application to a Regional Prediction of Advection Fog, Boundary-Layer Meteorology, 119, 397–407, https://doi.org/10.1007/s10546-005-9030-8, 2006.

NatLabRockies: Northing Calibration in FLASC v2.5: Homogenization with HOGER, https://natlabrockies.github.io/flasc/examples/01_raw_data_processing/03_northing_calibration_hoger.html#homegenization-with-hoger, fLASC documentation, accessed 2026-08-18, 2026.

NREL: FLORIS Wake Modeling and Wind Farm Controls Software, https://github.com/NREL/floris, 2025.

Nygaard, N. G., Steen, S. T., Poulsen, L., and Pedersen, J. G.: Modelling cluster wakes and wind farm blockage, Journal of Physics: Conference Series, 1618, 062 072, https://doi.org/10.1088/1742-6596/1618/6/062072, 2020.

Nygaard, N. G., Poulsen, L., Svensson, E., and Pedersen, J. G.: Large-scale benchmarking of wake models for offshore wind farms, Journal of Physics: Conference Series, 2265, 022 008, https://doi.org/10.1088/1742-6596/2265/2/022008, 2022.

O'Malley, M., Holttinen, H., Cutululis, N., Vrana, T. K., King, J., Gevorgian, V., Wang, X., Rajaei-Najafabadi, F., and Hadjileonidas, A.: Grand challenges of wind energy science – meeting the needs and services of the power system, Wind Energy Science, 9, 2087–2112, https://doi.org/10.5194/wes-9-2087-2024, 2024.

Palatos-Plexidas, A., Gremmo, S., Porchetta, S., and Munters, W.: A numerical analysis of wind farm wake characteristics in the southern part of the North Sea, Journal of Physics: Conference Series, 2767, 092 078, https://doi.org/10.1088/1742-6596/2767/9/092078, 2024.

Patel, A., Mann, J., Sjöholm, M., Zinck, K., and Raj, K.: Comparison of measured and simulated fatigue loads on a multi-megawatt wind turbine, Wind Energy Science, 11, 1147–1162, https://doi.org/10.5194/wes-11-1147-2026, 2026.

Pedersen, J. G., Svensson, E., Poulsen, L., and Nygaard, N. G.: Turbulence Optimized Park Model with Gaussian Wake Profile, Journal of Physics: Conference Series, 2265, 022 063, https://doi.org/10.1088/1742-6596/2265/2/022063, 2022.

Pedersen, M. M., Forsting, A. M., Van der Laan, P., Riva, R., Alcayaga, L., Criado Risco, J., Friis-Møller, M., Quick, J., Schøler Christiansen, J. P., Valotta Rodrigues, R., Olsen, B. T., and Réthoré, P.-E.: PyWake 2.5.0: An open-source wind farm simulation tool, https://gitlab.windenergy.dtu.dk/TOPFARM/PyWake, 2023.

Peña, A., Schaldemose Hansen, K., Ott, S., and Van Der Laan, M. P.: On Wake Modeling, Wind-Farm Gradients, and AEP Predictions at the Anholt Wind Farm, Wind Energy Science, 3, 191–202, https://doi.org/10.5194/wes-3-191-2018, 2018.

Platis, A., Siedersleben, S. K., Bange, J., Lampert, A., Bärfuss, K., Hankers, R., Cañadillas, B., Foreman, R., Schulz-Stellenfleth, J., Djath, B., Neumann, T., and Emeis, S.: First in situ evidence of wakes in the far field behind offshore wind farms, Scientific Reports, 8, 2163, https://doi.org/10.1038/s41598-018-20389-y, 2018.

Platis, A., Hundhausen, M., Lampert, A., Emeis, S., and Bange, J.: The Role of Atmospheric Stability and Turbulence in Offshore Wind-Farm Wakes in the German Bight, Boundary-Layer Meteorology, 182, 441–469, https://doi.org/10.1007/s10546-021-00668-4, 2022.

Porchetta, S., Munters, W., Lejeune, M., Borgers, R., Buckingham, S., and Howland, M. F.: Annual Wake Impacts in and between Wind Farm Clusters – Part 1: WRF-simulated Wake Losses for Different Atmospheric Conditions, Wind Energy Science, 11, 3273–3294, https://doi.org/10.5194/wes-11-3273-2026, 2026.

Quick, J., Mouradi, R.-S., Devesse, K., Mathieu, A., Paul Van Der Laan, M., Murcia Leon, J. P., and Schulte, J.: Verification and Validation of Wind Farm Flow Models, Journal of Physics: Conference Series, 2767, 092 074, https://doi.org/10.1088/1742-6596/2767/9/092074, 2024.

Quick, J., rsmouradi, Schulte, J., mathieu7an, DTUplaa, Olsen, B. T., and ernie: EUFLOW/WIFA: v0.1.0, https://doi.org/10.5281/zenodo.17058703, 2025.

Rosencrans, D., Lundquist, J. K., Optis, M., Rybchuk, A., Bodini, N., and Rossol, M.: Seasonal variability of wake impacts on US mid-Atlantic offshore wind plant power production, Wind Energy Science, 9, 555–583, https://doi.org/10.5194/wes-9-555-2024, 2024.

Schmidt, J., Vollmer, L., Dörenkämper, M., and Stoevesandt, B.: FOXES: Farm Optimization and eXtended Yield Evaluation Software, Journal of Open Source Software, 8, 5464, https://doi.org/10.21105/joss.05464, 2023.

Van Binsbergen, D., Daems, P.-J., Verstraeten, T., Nejad, A. R., and Helsen, J.: Hyperparameter Tuning Framework for Calibrating Analytical Wake Models Using SCADA Data of an Offshore Wind Farm, Wind Energy Science, 9, 1507–1526, https://doi.org/10.5194/wes-9-1507-2024, 2024.

van der Laan, M. P., Peña, A., Volker, P., Hansen, K. S., Sørensen, N. N., Ott, S., and Hasager, C. B.: Challenges in simulating coastal effects on an offshore wind farm, Journal of Physics: Conference Series, 854, 012 046, https://doi.org/10.1088/1742-6596/854/1/012046, 2017.

Vollmer, L., Wurps, H., and Dörenkämper, M.: Benchmarking engineering wake models for farm-to-farm interactions, Journal of Physics: Conference Series, 2767, 092 095, https://doi.org/10.1088/1742-6596/2767/9/092095, 2024.

wind-turbine-models.com: Siemens SWT-3.6-120 Onshore, https://en.wind-turbine-models.com/turbines/646-siemens-swt-3-6-120-onshore, accessed: 2026-05-21, 2026.

Zong, H. and Porté-Agel, F.: A momentum-conserving wake superposition method for wind farm power prediction, Journal of Fluid Mechanics, 889, A8, https://doi.org/10.1017/jfm.2020.77, 2020.