Wind Energ. Sci., 8, 747–770, 2023
https://doi.org/10.5194/wes-8-747-2023
© Author(s) 2023. This work is distributed under
the Creative Commons Attribution 4.0 License.

WIND ENERGY SCIENCE
Open Access

Validation of an interpretable data-driven wake model using lidar measurements from a field wake steering experiment

Balthazar Arnoldus Maria Sengers[1,2], Gerald Steinfeld[1], Paul Hulsman[1], and Martin Kühn[1]

[1]ForWind, Institute of Physics, Carl von Ossietzky Universität Oldenburg, Küpkersweg 70, 26129 Oldenburg, Germany
[2]current address: Fraunhofer IWES, Küpkersweg 70, 26129 Oldenburg, Germany

Correspondence: Balthazar Sengers (balthazar.sengers@uni-oldenburg.de)

Received: 15 December 2022 – Discussion started: 13 January 2023
Revised: 17 March 2023 – Accepted: 18 April 2023 – Published: 11 May 2023

Abstract. Data-driven wake models have recently shown a high accuracy in reproducing wake characteristics from numerical data sets. This study used wake measurements from a lidar-equipped commercial wind turbine and inflow measurements from a nearby meteorological mast to validate an interpretable data-driven surrogate wake model. The trained data-driven model was then compared to a state-of-the-art analytical wake model. A multi-plane lidar measurement strategy captured the occurrence of the wake curl during yaw misalignment, which had not yet conclusively been observed in the field. The comparison between the wake models showed that the available power estimations of a virtual turbine situated four rotor diameters downstream were significantly more accurate with the data-driven model than with the analytical model. The mean absolute percentage error was reduced by 19 % to 36 %, depending on the input variables used. Especially under turbine yaw misalignment and high vertical shear, the data-driven model performed better. Further analysis suggested that the accuracy of the data-driven model is hardly affected when using only supervisory control and data acquisition (SCADA) data as input. Although the results are only obtained for a single turbine type, downstream distance and range of yaw misalignments, the outcome of this study is believed to demonstrate the potential of data-driven wake models.

1 Introduction

With the wind energy industry maturing, more focus is being put on maximizing the power yield of existing assets. This involves moving away from the traditional, and currently still standard, greedy control of individual turbines to an optimization on the wind farm level. In recent years, especially the wake steering concept has received considerable attention in the literature; in this concept the turbine is intentionally misaligned with the inflow wind, introducing a lateral component of the thrust force that deflects the wake away from a downstream turbine. Many aspects of this strategy have been studied over the years, including the underlying physics (e.g., Howland et al., 2016; Bastankhah and Porté-Agel, 2016) and its characteristics under different atmospheric conditions (e.g., Vollmer et al., 2016; Schottler et al., 2017). Additionally, the implementation of this concept in the field with so-called yaw controllers has received attention. Such controllers typically include a representation of the wake in the form of engineering wake models used to solve the optimization problem, as well as the design of the yaw controller itself (e.g., wind direction robustness – Rott et al., 2018; Simley et al., 2020; hysteresis – Kanev, 2020; and open- versus closed-loop – Doekemeijer et al., 2020; Howland et al., 2020).

Although a large body of knowledge about the wake steering concept has been obtained, the industry appears to be hesitant to adapt due to the large uncertainties and lack of validation (van Wingerden et al., 2020; Boccolini et al., 2021). One limitation is the number of field experiments carried out. Due to the considerable expense and inaccessibility of test turbines, most research groups revert to high-fidelity simu-

lations or wind tunnel experiments. Although they provide a higher degree of reproducibility and more flexibility in choosing the studied scenarios, these experiments take place in controlled environments and do not fully represent the complexity of the field. Wake models and yaw controllers are consequently developed based on data from idealized conditions. Their accuracy in field situations is questionable due to limited validation, slowing down the adoption by industry. This uncertainty is amplified by findings that the application of wake steering can lead to power losses under certain conditions (e.g., Fleming et al., 2020; Doekemeijer et al., 2021).

Several field campaigns have been conducted in recent years to study wake steering control. In their pioneering work, Wagenaar et al. (2012) used a scaled wind farm to demonstrate the concept. Using rear-facing nacelle-mounted lidars, asymmetries in wake deflection depending on the sign of the yaw angle were observed for the near wake (Trujillo et al., 2016) and far wake (Bromm et al., 2018). This asymmetry is also found using numerical tools (e.g., Fleming et al., 2015) and attributed to shear-induced initial wake deflection (Gebraad et al., 2016) or the Coriolis force (Archer and Vasel-Be-Hagh, 2019). One prominent aspect associated with wake steering is the development of the wake curl as observed in numerical and wind tunnel experiments (e.g., Howland et al., 2016; Vollmer et al., 2016; Hulsman et al., 2022b). Fleming et al. (2017a) included a short notion that a curled shape could be observed in the field, while Brugger et al. (2020) did not find a curled wake in their field experiment. They argued that the effect of wind veer was too large for the counter-rotating vortices to generate a curled wake, with wind veer reported to tilt the wake in one direction (Herges et al., 2017; Brugger et al., 2019).

Using fixed yaw misalignment angles, Howland et al. (2019) found statistically significant gains of up to 47 % for low wind speeds and a certain wind direction in a small wind farm consisting of six turbines. Ahmad et al. (2019) reported that wake steering is mainly beneficial in partial wake situations. Fleming et al. (2021) found an asymmetry of the downstream turbine power generation, where gains from correct steering (wake steered away from turbine) are larger than the losses from erroneous steering (wake steered into turbine). They attributed this effect to the added wake recovery induced by the counter-rotating vortices that also generate the wake curl.

Additionally, several controller test studies have been carried out in which instead of a fixed yaw angle, an optimal yaw angle is employed based on the inflow conditions. This optimal yaw angle is determined with low-fidelity wake models which generate discretized lookup tables (LUTs). In a series of papers from the National Renewable Energy Laboratory (NREL), different versions of the FLOw Redirection and Induction in Steady State (FLORIS; NREL, 2022) framework have been used to generate these LUTs. In a field campaign at an offshore wind farm with a turbine spacing of seven to eight rotor diameters, Fleming et al. (2017b) reported a 10 % power gain for certain wind directions. Fleming et al. (2019, 2020) showed results of a field test with closely spaced turbines with two different versions of FLORIS, both resulting in a power gain for most conditions but clear power losses for some wind directions. Lastly, Doekemeijer et al. (2021) found large power gains of up to 35 % for one wind direction sector with a two-turbine setup in complex terrain, but also here large losses were found for other wind directions.

These studies are pivotal in demonstrating the potential of wake steering but also indicate that there is a large variability in its demonstrated effectiveness. Next to atmospheric inflow conditions, this can be attributed to turbine type, turbine spacing and terrain. Additionally, the choice of yaw controller and accuracy of the wake model used to develop the LUTs are believed to have an effect.

After the pioneering wake deficit models of Jensen (1983) and Ainslie (1988), Jimenez et al. (2010) first came up with a wake deflection model under yaw misalignment. Nowadays, most analytical wake models are based on the Gaussian model (Bastankhah and Porté-Agel, 2014, 2016; Niayifar and Porté-Agel, 2016). Combined with the curl wake model (Martínez-Tossas et al., 2019), the Gaussian–curl hybrid (GCH) model (King et al., 2021) prescribes the effect of counter-rotating vortices generated by turbine yaw misalignment, such as yaw-induced wake recovery, asymmetric deflection and secondary steering. Lastly, Bastankhah et al. (2022) presented an analytical way to describe the development of the wake curl with downstream distance, and Bay et al. (2023) tackled "deep array" effects, in which many wakes interact deep inside a large wind farm, with the cumulative-curl model.

In addition to these analytical models, data-driven wake (surrogate) models have received some attention in recent years. Most use complex neural networks (e.g., Ti et al., 2020; Renganathan et al., 2022; Purohit et al., 2022; Asmuth and Korb, 2022) and have shown highly accurate results. However, these models need lots of training data and have an extremely low interpretability (black-box models). In an attempt to overcome this, Sengers et al. (2022) presented an interpretable Data-driven wAke steeRing surrogaTe model (DART). Using only linear equations, DART uses inflow and turbine variables to estimate wake parameters such as deficit, center location and curl. It has a reduced number of parameters and is therefore highly interpretable and needs fewer training data. In a comparison using large-eddy simulation (LES) results, Sengers et al. (2022) demonstrated that DART outperformed the Gaussian and GCH models, especially under stable atmospheric conditions.

As mentioned before, studies validating wake models with field measurements are rare, especially when yaw misalignments are included, resulting in uncertainties about their accuracy. Moreover, comparisons between analytical and data-driven models in their abilities to reproduce the characteristics of wakes observed in the field are done sporadically.

However, validations with measurements and comparisons between models are necessary to assess their performance and provide direction for future work.

The objective of this paper is to use nacelle-based lidar measurements of the wake of a commercial turbine to validate the DART model and compare its accuracy with that of the GCH model. To achieve the objective, this study comprises three components: (1) to design a scanning strategy able to capture wake characteristics such as deficit, center position and curl to accurately reconstruct a vertical cross-section of the wake; (2) to assess the performance of the wake models by their ability to estimate the available power of a virtual downstream turbine observed by the lidar; and (3) to investigate DART's performance as a function of data set size and input variables, including an analysis of whether the model could operate on supervisory control and data acquisition (SCADA) data alone.

2 Measurement campaign

This section introduces the field experiment carried out within this study. Section 2.1 describes the measurement site and general setup. Section 2.2 describes the yaw control experiment. Section 2.3 through 2.7 then discuss the devices, their measurement strategies and data processing. Especially in Sect. 2.3 more details are provided, including results from a preliminary study to determine the scanning strategy of the nacelle lidar, since the measurements from this device are essential for this study. Lastly, Sect. 2.8 describes how the data from all devices are used to select 10 min averaged cases considered in the rest of the study.

2.1 Measurement site

Measurements were carried out in the period of February through April 2021 as a part of a yaw control field campaign at a slightly hilly onshore site in northeastern Germany located approximately 13.5 km from the Baltic Sea; see also Hulsman et al. (2022a). The layout of the site, including the positioning of the measurement equipment, is shown in Fig. 1. The nacelle of turbine T1 was equipped with a downstream-facing Leosphere WindCube 200S (serial no. WLS200S-024) pulsed lidar (Sect. 2.3). T1 was a commercial 3.5 MW eno126 turbine with a hub height of 117 m and a rotor diameter D of 126 m. The nacelle was further equipped with a Thies Clima wind vane and cup anemometer (Sect. 2.6), as well as a Trimble SPS three-antenna GNSS (hereafter called GPS) to measure orientation, tilt and roll (Sect. 2.7). A second pulsed lidar of the same type (serial no. WLS200S-023) was installed west of the turbine to measure inflow profiles (Sect. 2.4). North of this turbine, a meteorological mast (MM; Sect. 2.5) was erected and equipped with Thies Clima cup anemometers and wind vanes.

Lastly, Fig. 1 shows that a small 6 m high hill $5D$ upstream of T1 and a larger 27 m high hill $8D$ downstream of T1 were exactly in the wind direction sector that was not used due to the presence of the downstream turbines (see Sect. 2.2). Two villages with low buildings were located about 1 km from T1, directly upstream for wind directions around $\delta = 265$ and $\delta = 320°$, mainly outside of the studied wind direction sectors. The dominant vegetation in the area is of an agricultural nature, with patches of trees and bushes between the fields. These trees could affect the measurements for $\delta \approx 350°$, as noted in Hulsman et al. (2022a) using data from the same site. This influence was accepted, as omitting this sector would result in large data losses.

2.2 Yaw control experiment

As these measurements were part of a larger field campaign, only the wind direction sector $\delta = [268°, 360°] \cup [0°, 20°]$ could be used for experiments for this study. Unfortunately, in this sector two smaller turbines (T3 and T4) were located downstream of the lidar-equipped turbine. For the objectives of this study, measurements at $4D$ downstream were targeted. This was to avoid the near wake, as the two investigated wake models fail to represent the non-Gaussian shape of the wake deficit, and to ensure that the wake curl had developed. The wind speed reduction due to the induction zone of T3 at $4.8D$ (hub height of 103 m and a diameter of 93 m) was estimated to be on the order of 2 % (estimated with the vortex sheet theory – Medici et al., 2011) at the targeted distance of $4D$. Although not ideal, no alternative was possible due to the restrictions of the measurement site, and it was decided to neglect the effects of this induction zone.

Part of the wind direction sector could not be used due to the positioning of T4 at $3.2D$ downstream. To make sure that the wake was not steered into T4, in the sector $\delta = [268, 316°]$ the turbine toggled between target yaw misalignment angles of $\phi_t = 0°$ (duration of 30 min) and $\phi_t = +15°$ (duration of 60 min, clockwise rotation looking from above), steering the wake to the left. Correspondingly, in the sector $\delta = [316°, 360°] \cup [0°, 20°]$ the turbine toggled between $\phi_t = 0°$ (30 min) and $\phi_t = -15°$ (60 min, counterclockwise rotation looking from above), steering the wake to the right. The downside of this approach was that directly comparing positive and negative yaw angles under similar atmospheric conditions was not possible. Additionally, more data were collected in the first sector as this wind direction was more dominant.

Fixed yaw offsets were applied as this involved minimal changes to the yaw controller. Besides, a distribution of yaw misalignments was expected to be obtained due to the imperfect tracking of the wind direction by the yaw controller.

2.3 Nacelle lidar

This subsection describes the measurements performed with the nacelle-mounted lidar. Section 2.3.1 describes the design of the scanning strategy, including results of a numer-

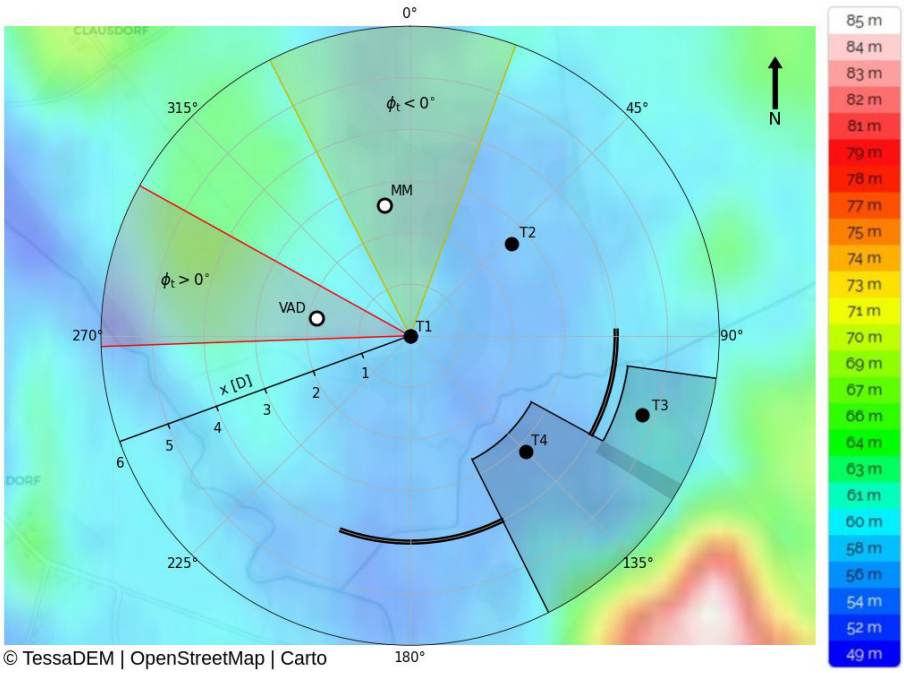

Figure 1. Layout of the measurement site with the local topography, relative to mean sea level, indicated in the background. Black markers indicate turbines, where T1 is equipped with the nacelle lidar. White markers indicate the meteorological mast (MM) and ground-based lidar (VAD). Shaded areas indicate the wind direction sector with $\phi_t > 0°$ (red) and $\phi_t < 0°$ (yellow) and where wake measurements are assumed to be disturbed by the downstream turbines (grey). The thick solid black line indicates the measured locations used for analysis. (Source for topographic map including color bar: topographic-map.com, 2022.)

ical evaluation to determine what trajectory should be implemented in the field. Section 2.3.2 describes the processing, including filtering, of these data.

2.3.1 Design scanning strategy

A pulsed lidar can be mounted onto the nacelle to sample to turbine's wake. When operated with a single plan position indicator (PPI) scan with an elevation angle of $\phi_{PPI} = 0°$, the line-of-sight velocities on a horizontal plane at hub height are obtained. Although quick, this trajectory only provides data at one height in the wake. Attempts have been made to capture information in the vertical plane, such as in Beck and Kühn (2019), who proposed a scanning pattern of alternating PPI and range height indicator (RHI) scans to obtain information in both dimensions. However, wake shape deformations due to wind veer (tilted) or yaw misalignment (curled) cannot be captured with this scanning strategy. Brugger et al. (2019, 2020) used nine PPI scans at different elevation angles, allowing the description of non-circular wake shapes in a vertical plane.

In this paper, their strategy was adopted and evaluated numerically to gain insights into how the number of PPI scans and their angular speed (following Carbajo Fuertes and Porté-Agel, 2018) affect the ability to capture the characteristics of 10 min averaged wake. This exercise used large-eddy simulation (LES) results, allowing for a systematic uncertainty analysis of the proposed scanning patterns.

The Parallelized Large-Eddy Simulation Model (PALM; Maronga et al., 2020) coupled with the aeroelastic code FAST (Jonkman and Buhl, 2005; Krüger et al., 2022) representing the NREL 5 MW turbine (Jonkman et al., 2009) provides the numerical wind fields. Precursor simulations generated realistic inflow conditions, after which the main simulations with one turbine were performed. The aeroelastic code for the turbine installed in the field, as used in Sect. 4.1, was not yet available during the planning stage of this campaign. Both turbine T1 and the NREL 5 MW turbine have the same rotor diameter (126 m) but differ in hub height (117 vs. 90 m) and aeroelastic properties. It was, however, assumed that at $4D$ the characteristics of the wakes produced by these turbines are sufficiently similar.

A single turbine with yaw angles of $\phi = \{-15, 0, 15°\}$ in a neutral (TI = 10.3 %, $\alpha = 0.17$; see the Appendix for definitions of abbreviations and symbols used throughout) and a stable (TI = 5.7 % and $\alpha = 0.32$) boundary layer with a hub height wind speed $U_h \approx 8\,\mathrm{m\,s^{-1}}$ was simulated. The simulation length was 25 min, of which the first 15 min was omitted as spin-up and the remaining 10 min was used for analysis. Synthetic lidar data targeting $4D$ downstream were subsequently generated by employing the lidar simulator LiXim (Trabucchi, 2019) with an accumulation time of 0.1 s and an

opening angle of 70°. Temporal averages were taken for all points in the scanning cycle. The wake composition method, later described in Sect. 3.1, was used to reconstruct vertical cross-sections of the wake, from which the available power P_{av} could be determined. This estimate is compared to the reference (subscript ref) 10 min averaged LES data. Used as a metric is the absolute percentage error (APE) over the six (two boundary layers times three yaw angles) simulations calculated with Eq. (1):

$$\mathrm{APE}[\%] = \left| \frac{P_{\mathrm{av}} - P_{\mathrm{av,ref}}}{P_{\mathrm{av,ref}}} \right| \cdot 100, \tag{1}$$

in which $P_{\mathrm{av}} = P/C_{\mathrm{P}} = 0.5\rho A U_{\mathrm{eq}}$ with ρ the air density (assumed to be constant), A the rotor area and U_{eq} the rotor-equivalent wind speed. The bar "all" on the far left in Fig. 2 indicates the reconstruction of the wake based on the original LES data, hence the error introduced by the composition method. Further, one, three, five, seven and nine PPI scans were tested, where the middle scans always targeted hub height and the outermost scans upper and lower tip height at $4D$. Trajectories with an even number of PPI scans were not tested, as this would remove the scan at hub height that was needed for another study. Additionally, it is desirable to measure the largest wake deficit, which is expected to develop around hub height.

Figure 2 shows that five PPI scans typically hold the highest accuracy. Using fewer PPI scans results in inaccurate estimations of the wake deficit distribution in the vertical, while using more PPI scans results in long cycles and consequently fewer measurements per observation point. The angular speed ω_{lidar} seems to have little effect, except for when seven PPI scans are used. This is attributed to chance, as too few cases are studied for the statistics to converge. Generating more LES results with a wider range of atmospheric conditions and turbine yaw angles was not possible due to computational restrictions. While these results are not statistically significant and it can therefore not be claimed that an "optimal" scanning strategy has been found, this exercise allows for making an informed decision.

It was decided to implement the trajectory showing the lowest error, hence consisting of five PPI scans with $\omega_{\mathrm{lidar}} = 14°\,\mathrm{s}^{-1}$. The elevation angles of these scans were $\phi_{\mathrm{PPI}} = \{-7.0, -3.5, 0.0, 3.5, 7.0°\}$, and the accumulation time used was 0.1 s. With an opening angle of 70°, the duration of one PPI scan is 5 s. Changing elevation angles takes 1.3 s, and resetting to the start of the cycle takes 3.5 s, adding to 34 s to complete one full cycle. The range gate length was set to 25 m, corresponding to a pulse duration of 100 ns. Range gates were defined between 50 and 1340 m with 5 m spacing. However, in the processing phase only data up to 820 m were used to avoid the influence of the ground in the PPI scan with the lowest elevation angle.

2.3.2 Data processing

Since the performed PPI scans were quite fast with a relatively coarse resolution, all PPI scans with the same elevation angle in a 10 min window (see Sect. 2.8) were grouped together to get a better estimate of the measurement distribution.

Simple filtering based on the carrier-to-noise ratio (CNR) and line-of-sight velocity (LOS) was performed, where only realistic data with $\mathrm{CNR} < 0\,\mathrm{dB}$ and $0\,\mathrm{m\,s^{-1}} < \mathrm{LOS} < 20\,\mathrm{m\,s^{-1}}$ were kept. On the remainder, a Gaussian filter was used, retaining only measurements within 3 standard deviations of the median CNR and LOS (99 % confidence interval). This removed outliers due to hard targets, as illustrated in Fig. 3a.

However, some PPI scans exhibited a LOS–CNR diagram as illustrated in Fig. 3b, containing many measurements with high CNR and low LOS values. To filter out these erroneous measurements, a mean shift clustering algorithm (Fukunaga and Hostetler, 1975) was employed, as was for instance used in Wang et al. (2022) as part of data cleaning for power curve tuning. The algorithm identified clusters in the LOS–CNR space and allocated all measurements to any of the clusters based on the Euclidean distance to the cluster center. Clusters were then either considered or eliminated based on whether the location of their center was physically feasible. In the example in Fig. 3b, the yellow cluster was omitted, since many points outside the main cluster with high CNR and low LOS values indicate erroneous measurements. Lastly, the Gaussian filter based on the 99 % confidence interval was repeated, as removing one cluster drastically affected the outcome of this filter.

After filtering, all PPI scans were interpolated to a standard grid with a resolution of 1.4° (corresponding to the original resolution) to account for the slightly different azimuth angles between scans as a result of the lidar's inability to measure the exact same location each time. Next, the PPI scans were temporally averaged as long as not more than two data points within a 10 min window were missing. When more than 25 % of the measurements were filtered out, as is the case with Fig. 3b, the averaged PPI scan was removed from the 10 min window, resulting in fewer than five PPI scans. If fewer than four averaged PPI scans remained after filtering, the case was eliminated.

Lastly, the PPI scans' azimuth and elevation angles were corrected with the nacelle's 10 min averaged tilt angle and misalignment (see Sect. 2.7). The horizontal wind speed was subsequently computed by correcting the LOS with these azimuth and elevation angles.

2.4 Ground-based lidar (VAD)

As shown in Fig. 1, the ground-based lidar was situated $1.85D$ upstream of the lidar-equipped turbine for $\delta = 281°$ to measure profiles of wind speed and direction. The ground-

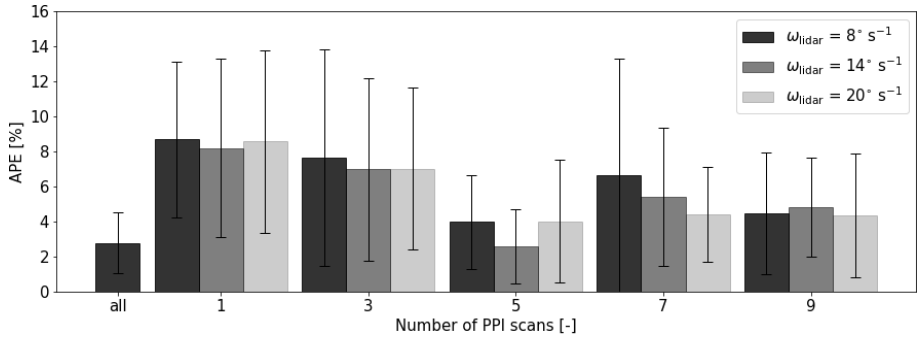

Figure 2. Results of the virtual lidar tests. Bars indicate the mean and whiskers the standard deviation of the absolute percentage error (APE) of available power P_{av} over six simulations. The number of PPI scans is indicated on the x axis. "all" indicates the use of all numerical data, hence the error introduced by the composition method. The opaqueness represents the lidar's angular speed ω_{lidar}.

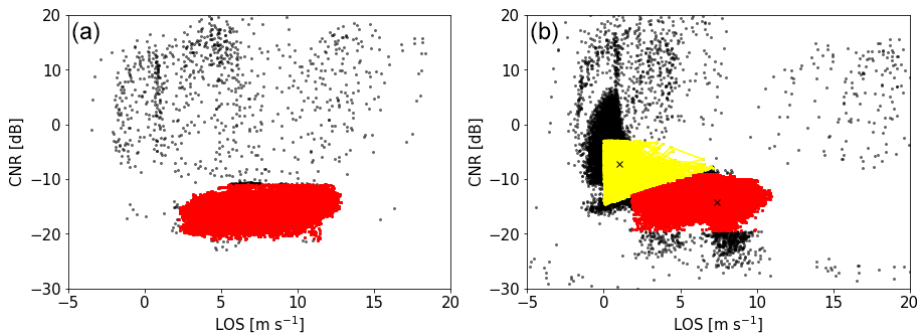

Figure 3. Examples of multiple PPI scan filtering in LOS–CNR diagrams in which black markers indicate original data and red markers data kept after filtering. **(a)** A textbook case with few outliers that indicate hard targets and **(b)** a more problematic case in which there are many corrupted measurements. Here yellow markers indicate a second cluster from which all measurements were omitted. Black crosses indicate the two cluster centers.

based lidar performs continuous velocity–azimuth display (VAD) scans at an elevation angle of $\phi_{VAD} = 75°$ with an accumulation time of 0.5 s and an angular speed of 30° s^{-1}. Also for this lidar, the range gate length was set to 25 m, corresponding to a pulse duration of 100 ns. Range gates were defined between 50 and 840 m with 5 m spacing.

Filtering was done based on the 2D histogram method introduced by Beck and Kühn (2017), which assumes a normal distribution of LOS and CNR values. The measured data points were binned by their LOS and CNR values, and the number of data points in each bin were counted. Bins having a count below 10 % of the bin with the highest count were omitted.

Next, the azimuth angle (θ_{VAD}) was corrected by means of a hard-target analysis, such that $\theta_{VAD} = 0°$ faces north. To obtain the wind speed components (u, v, w) and consequently the horizontal wind speed and direction, the measurements of each range gate were fitted with the following sinusoid:

$$\mathrm{LOS} = u\cos(\theta_{VAD})\sin\left(\frac{\pi}{2} - \phi_{VAD}\right) + v\sin(\theta_{VAD})\sin\left(\frac{\pi}{2} - \phi_{VAD}\right) + w\cos\left(\frac{\pi}{2} - \phi_{VAD}\right). \quad (2)$$

Lastly, only when at least 75 % of the data points remained after filtering and the fitted sinusoid achieved a correlation coefficient of at least 0.8 (determined empirically), the wind speed components of a vertical level were retained.

2.5 Met mast

A meteorological (met) mast was positioned 2.7D upstream of T1 at $\delta = 350°$ (Fig. 1). This mast was equipped with cup anemometers at 116.3 m (hub height, U_h) and 54.2 m (lower tip height, U_{lt}) to measure wind speed and shear. Wind vanes were located at 112.2 m (approximately hub height, δ_h) and 54.5 m (lower tip height, δ_{lt}). The highest cup anemometer was located on the top of the met mast for undisturbed flow from all directions, whereas the other cup anemometer and wind vanes had orientations of 315 and 135°, respectively. A flow distortion due to the tower structure affecting the measurements occurs for wind directions between approximately 310 and 320°, which is not considered in this study (see Sect. 2.1). The wind directions analyzed here are assumed to be undisturbed. The cup anemometers and vanes had an accuracy of 0.2 m s^{-1} and 1.5°, respectively. All sensors operated at a sampling frequency of 50 Hz.

2.6 Wind turbine operational data

SCADA data were collected at the turbine at a frequency of 50 Hz. These data contain measurements from the nacelle's wind vane δ_S and cup anemometer U_S, as well as power P, rotor speed ω and turbine status, the latter indicating whether the turbine was operating normally. A standard nacelle transfer function was used by the operator to correct wind speed measurements for the influence of the rotor.

2.7 GPS

All above-mentioned systems were equipped with a Global Positioning System (GPS) sensor used for time synchronization. Additionally, the nacelle of T1 was equipped with a three-antenna global navigation satellite system (GNSS) to measure orientation, roll and tilt. This system was operated at a sampling frequency of 10 Hz, and its measurements have a root mean square error of less than 0.1°. This results in a spatial error of less than 1 m at $4D$ downstream.

Orientation measurements, averaged to 10 min values to smooth out high-frequency vibrations, were used to compute the yaw misalignment ϕ of the turbine relative to the wind direction δ_h measured at the met mast. These measurements were then used to correct the PPI scans' azimuth angles. Likewise, 10 min averaged nacelle tilt angles were used to correct the PPI scans' elevation angles, but the scans were not corrected for roll as it was expected to only have a small influence on the results.

2.8 Selection of data for model evaluation

The measurements were averaged over 10 min as is commonly done in the wind energy industry. Case selection was done using the following steps:

1. Within a 10 min window, no yaw maneuver should take place. A preselection of cases was therefore done purely based on GPS data. A case was considered when the orientation did not change for at least 12 min, of which the first 2 min was not considered for analysis because the wake needed time to reach $4D$ downstream. In the case that the orientation did not change for more than 22 min, the first 2 min was omitted and the remainder was split into two 10 min windows as far apart as possible.

2. The 10 min averaged U_h needed to be between cut-in and rated wind speed. δ_h needed to be in the defined sector (Sect. 2.1) and approximately normally distributed. This eliminated situations where there is a clear trend in the wind direction signal.

3. The inflow measured at the met mast should reasonably compare to the measurements at the turbine's nacelle. The met mast measurements were temporally corrected to match the nacelle signal using Taylor's hypothesis of frozen turbulence. Next, the two signals were compared, where the 10 min averaged wind speed $|U_h - U_S| < 1\,\mathrm{m\,s^{-1}}$ and direction $|\delta_h - \delta_S| < 5°$.

4. The profiles from the VAD lidar were used to check whether the wind speed profiles were approximately logarithmic, as the effect of low-level jets on the downstream wake characteristics is currently not captured by the wake models and considered out of the scope of this study.

5. If all checks were passed, all completed cycles within the defined 10 min window were averaged as described in Sect. 2.3. After averaging, the PPI scans were interpolated to a vertical plane at $4D$ downstream of the turbine. The wake deficit (U_{def}) was calculated by subtracting the wake measurements with the inflow profile obtained from the met mast measurements and normalized by dividing by the hub height wind speed U_h.

6. Lastly, the 10 min averaged cases were evaluated by the multiple 1D Gaussian method (see Sect. 3.1). Since the opening angle of the PPI scans is 70°, it can be expected that wakes from other turbines are also visible in the measurements. To prevent using an incorrect wake, the scans are sliced around the expected location of the considered wake. Boundaries of these slices are determined by the maximum wind speeds between the scan's center, corrected for yaw misalignment, and 150 m left and right of this center. Furthermore, the correlation coefficient (R) of the Gaussian fit with the wake deficit observations needed to be higher than 0.85 (empirically determined) to be considered, removing cases that do not fulfill the model assumptions of a Gaussian wake deficit.

This selection procedure resulted in 382 individual 10 min averaged cases to be used for analysis. Figure 4 displays the distribution of measured yaw angles during the campaign. Most measurements were done without yaw misalignment, since during a part of the campaign the implemented controller had issues and turbine control reverted back to standard operation. The difference between the number of positive and negative yaw angles is due to a more dominant wind direction in the sector containing positive yaw angles.

The solid vertical lines indicate the median yaw angles per target angle. For greedy control, the median shows a small bias of $\phi = -0.94°$, suggesting a calibration error of the nacelle's wind vane. For a target angle $\phi_t = +15°$, the median achieved $\phi = +11.14°$, whereas for $\phi_t = -15°$, $\phi = -13.19°$ is achieved. These angles are smaller than the targeted angles, which is due to the wind vane error under yaw misalignment (Kragh and Fleming, 2012; Simley et al., 2021a). Figure 5 displays an overview of the inflow conditions measured during these 382 cases. The shear α with a mean of 0.3 is slightly larger than expected, and the veer $\delta\alpha$ is smaller than expected, showing a high occurrence of neg-

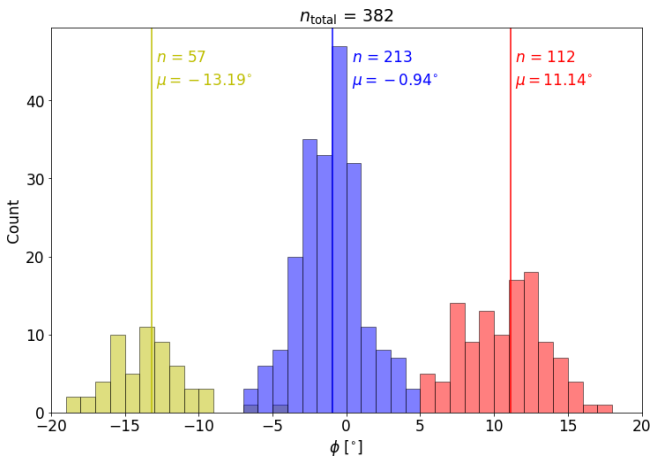

Figure 4. Data availability of the 10 min averaged cases as a function of achieved yaw angle (ϕ). Colors indicate the targeted $\phi_t = -15°$ (yellow), $\phi_t = 0°$ (blue) and $\phi_t = +15°$ (red). Solid vertical lines and accompanying text mark the median of the achieved yaw angles.

ative values. Regardless, all variables show a range of values that are physically reasonable.

3 Methods

This section introduces the modeling aspects of this study. First, Sect. 3.1 summarizes the multiple 1D Gaussian method used to obtain quantifiable wake characteristics. Section 3.2 discusses what information is used as a reference, and Sect. 3.3 describes the splitting of the data set into training and testing subsets. Then, Sect. 3.4 introduces the data-driven model and Sect. 3.5 briefly introduces the analytical model used in this study.

3.1 Multiple 1D Gaussian method

The multiple 1D Gaussian method (Sengers et al., 2020) is utilized to obtain quantifiable wake characteristics, listed in Table 1. This method fits a 1D Gaussian through the wake deficit data normalized by the wind speed at hub height (U_{def}/U_h) in the horizontal plane for every height level, in the current study obtained from five consecutive PPI scans. This results in a set of local wake deficits (amplitude), center positions (location) and widths (standard deviation) for each height. By fitting another 1D Gaussian through the set of local deficits in the vertical, the vertical deficit profile can be determined. The position of the maximum deficit in this profile is then considered the vertical position of the wake center. The horizontal position of the wake center is determined by interpolating the set of local center positions to this height. A second-order polynomial is fit through the set of local wake center positions to find the wake curl and tilt. The same method is applied for the wake widths to find their profile as a function of height.

Table 1. Dimensionless variables describing the wake characteristics obtained with the multiple 1D Gaussian method. Reused from Sengers et al. (2022) with permission.

Scalar parameter	Symbol
Amplitude normalized wake deficit	A_z
Lateral wake center displacement	μ_y
Vertical wake center displacement	μ_z
Width wake center height	σ_y
Vertical extent	σ_z
Curl	curl
Tilt	tilt
Quadratic wake width parameter	s_a
Linear wake width parameter	s_b

The reverse of this method, hereafter called the composition method, can be used to obtain a vertical cross-section of the wake from a set of wake characteristics. For more details on the multiple 1D Gaussian method and the composition method, the reader is referred to Sengers et al. (2020, 2022).

3.2 Reference power

The wind speed measured by the nacelle lidar is used to obtain the available power at $4D$ downstream. Since the spatial resolution is relatively coarse and the two outermost PPI scans target the tip heights, the 10 min averaged wind speeds are interpolated using a cubic spline function to a resolution of $\Delta = 5$ m. This inherently fills gaps when data are not available. The spatially interpolated data are consequently used to determine a rotor-equivalent wind speed (U_{eq}) and an available power $P_{av,ref}$ used as a reference in the remainder of this study.

3.3 Training and testing data

The data set is split into a training part (80 % of total size) and a testing part (remaining 20 %). This has been done in a stratified random manner, meaning the data set was first split up into three subsets according to their target yaw angle $\phi_t = \{-15, 0, 15°\}$, after which from each subset 20 % of data were randomly selected to be testing data. This way, it is ensured that both testing and training data contain cases with a yaw misalignment.

To not base the results on only a single testing data set, this random splitting of the data set (resampling) has been repeated 96 times (hereafter: resamples). The choice of 96 resamples is pragmatic, as it was convenient for parallel computing (multiple of 24 nodes per core). Error statistics appear to be normally distributed, which was not the case with 24, 48 or 72 resamples. Although more resamples are desirable (e.g., bootstrapping is typically done over several thousands),

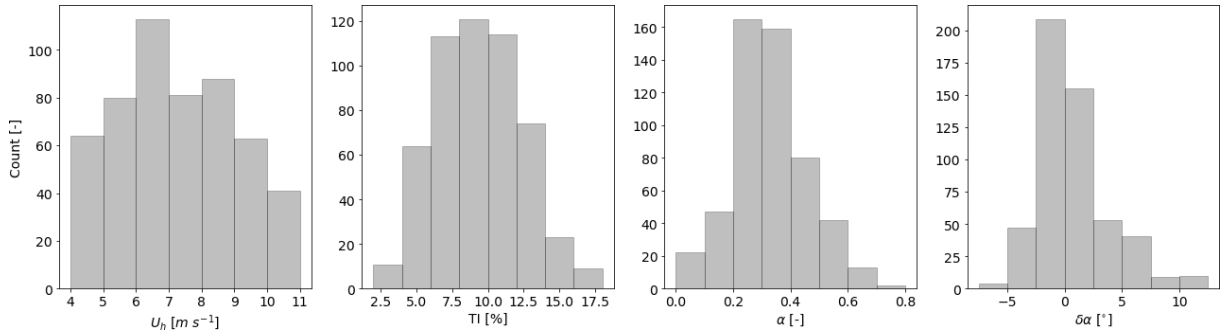

Figure 5. Distribution of 10 min averaged inflow variables measured at the met mast for all 382 cases: hub height wind speed (U_h), turbulence intensity (TI), shear (α) and veer ($\delta\alpha$).

this was not possible due to computational limitations as the training of the models can be quite expensive as discussed in Sect. 3.4.4 and 3.5.

3.4 Data-driven wAke steeRing surrogaTe model (DART)

This subsection introduces DART, starting with a summary from previous work in Sect. 3.4.1 and changes made to the model since this work in Sect. 3.4.2. This is followed by information on the input variables (Sect. 3.4.3). Lastly, the feature selection of the three versions of the model considered in this study is discussed in Sect. 3.4.4.

3.4.1 Model description

DART was introduced in an LES study by Sengers et al. (2022). It estimates wake characteristics (Table 1) obtained from the multiple 1D Gaussian method (Sect. 3.1) with a linear regression model from standard input parameters (e.g., yaw misalignment, shear, thrust coefficient). These wake characteristics ($\boldsymbol{Y}$) are estimated from input parameters ($\mathbf{X}$) using a simple linear model:

$$\underset{(n)}{\boldsymbol{Y}} = \underset{(n\times p)}{\mathbf{X}} \times \underset{(p)}{\boldsymbol{B}}, \tag{3}$$

in which $\boldsymbol{B}$ denotes the model coefficients. The matrix dimensions are indicated by the sample size n and the number of input parameters p, containing the input variables, their second-order and interaction terms, and intercepts. The model coefficients are fitted with the lasso method (Tibshirani, 1996), using the following cost function:

$$\underset{B}{\operatorname{argmin}} \sum_n (y_n - \sum_p x_{np}\boldsymbol{B}_p)^2 + \lambda \sum_p |\boldsymbol{B}_p|. \tag{4}$$

This method remains close to ordinary least squares but adds a regularization parameter λ to its cost function, effectively penalizing adding more parameters. This ensures shrinkage of the input parameters and eliminates the issue of multicollinearity as only one of the highly correlated input parameters is chosen. The notations presented here deviate slightly from those in Sengers et al. (2022), as in the current study only one distance downstream is considered, simplifying the equations.

To include nonlinear relations between input parameters and wake characteristics, the original variables can be transformed with, e.g., a square root or exponential transformation. In the training stage (Sect. 3.4.4), it is determined what set of input variables and transformations yields the most accurate results.

Lastly, the estimated wake characteristics are used in a composition method (Sect. 3.1) to generate a vertical cross-section of the wake deficit and wind field. For a more detailed description of DART, the reader is referred to Sengers et al. (2022).

3.4.2 Modifications to the model

A few changes have been made to DART since its first description in Sengers et al. (2022). Most notably, the feature selection procedure has been changed. Before estimating the wake characteristics with a linear model, inflow variables (e.g., ϕ, α, ω) undergo transformations. In addition to the non-transformed variable, the square root, exponent, natural logarithm and reciprocal transformation are considered for all input variables, resulting in five options for each variable and many possible sets of input parameters (e.g., $\phi, \alpha^{-1}, \ln(\omega)$). In Sengers et al. (2022), all these possibilities were tested, the available power of a virtual turbine was estimated and compared to the original data, and the set of input parameters that had the smallest error was chosen. This selection procedure not only was very computationally expensive but also does not necessarily give the most accurate solution for all wake characteristics. Hypothetically, the wake center position could be best explained by the non-transformed yaw angle, whereas the wake curl could be best explained by the exponent of the yaw angle. In the current work, the determination of the best set of transformations is tested for each wake steering variable individually. The best transformation is then chosen as the one that has the smallest mean abso-

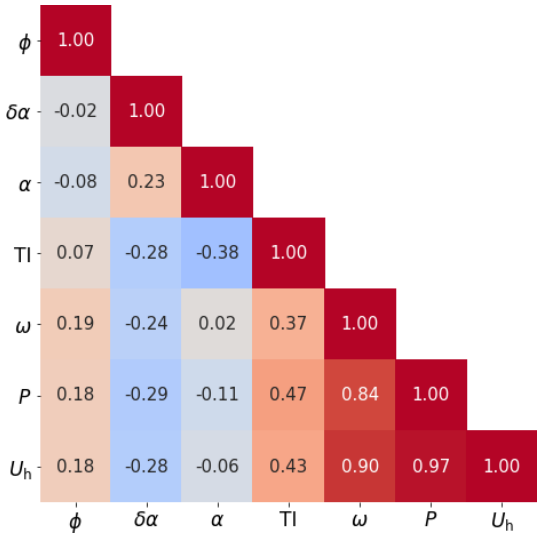

Figure 6. Correlation matrix of available input variables. In addition to the inflow variables shown in Fig. 5, yaw misalignment ϕ, rotor speed ω and power P are considered input variables.

lute error in the training data. This not only allows for more accurate estimates but also speeds up the training process.

Secondly, square root and natural logarithm transformations do not allow for negative input values. A sign function is used to include these values rather than omitting them, as was done in the previous work.

Lastly, in the testing phase, extrapolation is prevented by using the maximum (or minimum) value found in the training data when an input variables exceeds this range. Although this does not allow DART to give accurate estimations in new situations, it eliminates erroneous estimates due to extrapolation.

3.4.3 Input variables

As argued in Sengers et al. (2022), highly correlated input variables are interchangeable as they provide similar information. However, as long as they are not perfectly correlated, including all variables can lead to a higher accuracy as some new information is added. Due to the use of the lasso regression method, multicollinearity is not an issue.

Because of DART's flexibility, training with different sets of input variables is possible, allowing for an analysis of the model's accuracy as a function of chosen input variables. An overview of the available input variables is displayed in a correlation matrix (using the Pearson correlation coefficient) in Fig. 6. Other variables such as the wind direction variability and TI at lower tip height could have been included but were omitted for brevity. As opposed to what was seen in LES in Sengers et al. (2022), the inflow variables $\delta\alpha$, α and TI are weakly correlated in this field experiment. Secondly, ω and P are highly correlated with U_h.

3.4.4 Feature selection

In this study, multiple versions of the DART model were considered, each having a different set of input variables. Adding more input variables might increase the accuracy but will increase the training time of the model significantly. In Sengers et al. (2022) it was hypothesized that DART can achieve reasonable accuracy as long as each of the following clusters is represented: yaw (ϕ), atmospheric inflow ($\delta\alpha$, α, TI) and turbine (ω, P, U_h). Due to its high correlation with the turbine variables, U_h is here considered a turbine variable rather than an inflow variable. Following this logic, the first version of DART uses three input variables.

DART-3

To determine the most accurate solution using only three variables, all possible sets of input variables and their respective transformations (Sect. 3.4.2) are tested during the training stage and their accuracy to reproduce the training data set is investigated. By means of an example, Fig. 7a displays the error distribution of one resample. The error metric used here is the percentage error (PE) of P_{av} (calculated analogously to Eq. 1 but without absolute values) at $4D$ downstream. From these values, a mean percentage error (MPE) and mean absolute percentage error (MAPE) can be computed, as indicated in the top right of the figure.

Repeating this for all 96 resamples, one can obtain a histogram of MPEs and MAPEs as displayed in Fig. 7b and c. Finally, the mean over 96 MPEs ($\overline{\text{MPE}}$) and MAPEs ($\overline{\text{MAPE}}$) can be calculated; see top right of the figures. The $\overline{\text{MPE}} = -2.19\,\%$ illustrates a negative systematic bias, meaning DART underestimates P_{av}.

To determine the most accurate set of input variables, $\overline{\text{MAPE}}$ is considered. The results for all considered combinations of input variables are displayed in Table 2, showing that the set $\{\phi, \alpha, P\}$ provides the most accurate result (lowest $\overline{\text{MAPE}}$) and is therefore used in the remainder of the study, denoted as DART-3. The training time for each set of input variables with DART-3 is on the order of 10 min; hence the total computation time to determine the best set of input variables is approximately 1.5 h.

DART-4

Because the training of DART-3 is fast, an additional variable can be included to improve the accuracy of the model. The first three variables are chosen similarly to DART-3 (one from each cluster), while the fourth variable can be any input variable not yet selected. Repeating the analysis of computing a MAPE for each resample and consequently a $\overline{\text{MAPE}}$ for each set of input variables, generating a table corresponding to Table 2 (not shown here for brevity), reveals that $\{\phi, \alpha, P, U_h\}$ is the most accurate combination with $\overline{\text{MAPE}} = 12.69\,\%$, hereafter called DART-4. Its training time for each combination is approximately 1 h; hence with

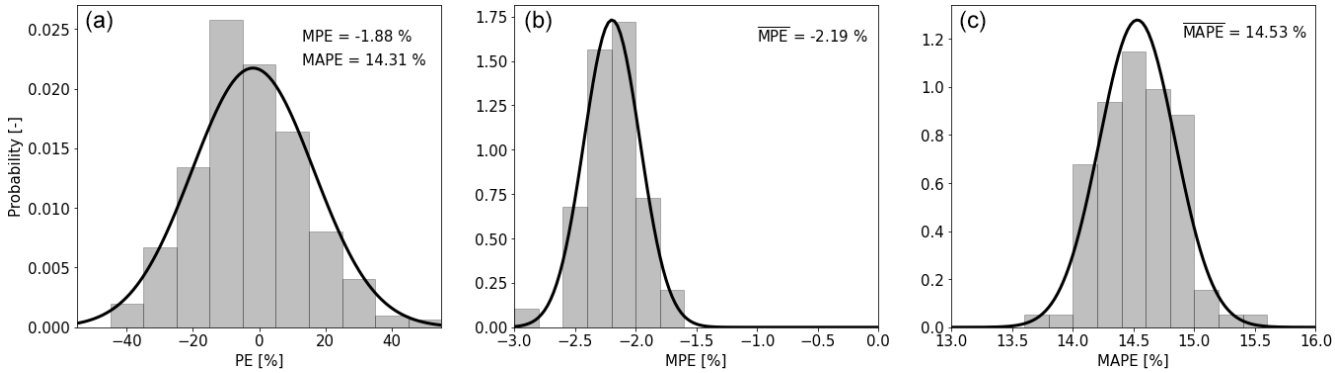

Figure 7. Performance of DART-3 on the training data. The set of input variables is $\{\phi, \alpha, P\}$. **(a)** Histogram of PE of P_{av} for one resample. Fitted normal distributions are indicated with solid lines, and MPE and MAPE are given in the top right. Histograms of MPEs **(b)** and MAPEs **(c)** over all 96 resamples.

Table 2. Overview of all possible combinations of input variables in DART-3 and their respective $\overline{\text{MAPE}}$ values. Boldface indicates the combination resulting in the lowest error.

Variable 1	Variable 2	Variable 3	$\overline{\text{MAPE}}$ [%]
ϕ	$\delta\alpha$	ω	18.24
ϕ	$\delta\alpha$	P	16.70
ϕ	$\delta\alpha$	U_{h}	17.17
ϕ	α	ω	15.11
$\boldsymbol{\phi}$	$\boldsymbol{\alpha}$	$\boldsymbol{P}$	**14.53**
ϕ	α	U_{h}	15.08
ϕ	TI	ω	15.66
ϕ	TI	P	14.77
ϕ	TI	U_{h}	14.81

18 possible sets of input variables, the total computation time needed for training is 18 h.

DART-7

Lastly, all available variables are used as input in DART-7, demonstrating the maximum achievable accuracy of the data-driven model during this experiment. DART-7's accuracy on the training data was indeed the highest with $\overline{\text{MAPE}} = 10.31\,\%$. The computation time needed to train DART-7 is approximately 1 month if not parallelized.

3.5 Analytical wake model

The state-of-the-art GCH model (King et al., 2021) as available in version 3.0rc4 of the FLORIS framework (NREL, 2022) acts as a reference model in this study. The GCH model incorporates the spanwise and vertical velocity components (Martínez-Tossas et al., 2019) due to the present vortices into the Gaussian wake model (Bastankhah and Porté-Agel, 2014, 2016; Niayifar and Porté-Agel, 2016).

Since presently only the wake at a distance of $4D$ behind the upstream turbine is studied, the GCH model could have benefited from including a near-wake model (e.g., Blondel and Cathelain, 2020), but this coupling was not available in this version of FLORIS. The C_{T} curve of the eno126 turbine is obtained from the Bladed model for which the aerodynamic properties of the turbine were provided by the operator and is used in these calculations.

Inflow information is taken from the 10 min averaged met mast data. The model tuning parameters (α_{GCH}; β_{GCH} for the far-wake onset; and $k_{a,\text{GCH}}$, $k_{b,\text{GCH}}$ for the wake growth rate) are determined by minimizing the MAPE of available power over the training data, analogously to the training of DART described in Sect. 3.4.4. The tuning takes about 3 h, and the model has an error of $\overline{\text{MAPE}} = 18.13\,\%$.

4 Results

This section presents the results of this study. Section 4.1 describes the characteristics of the wake observed in the field, after which in Sect. 4.2 the performance of the wake models in reproducing these wake characteristics is discussed.

4.1 Observed wake characteristics

In Sect. 4.1.1 an assessment of the characteristics of the observed wake listed in Table 1 is performed, which is deemed a necessary first step before investigating the accuracy of wake models. The observed wake characteristics are linked to the inflow variables to examine whether the measurements are physically feasible. In Sect. 4.1.2, two wake characteristics that are deemed important for wake steering are further investigated.

4.1.1 Correlation with inflow variables

The multiple 1D Gaussian method (Sect. 3.1) is used to describe the wake in quantifiable characteristics. Figure 8 dis-

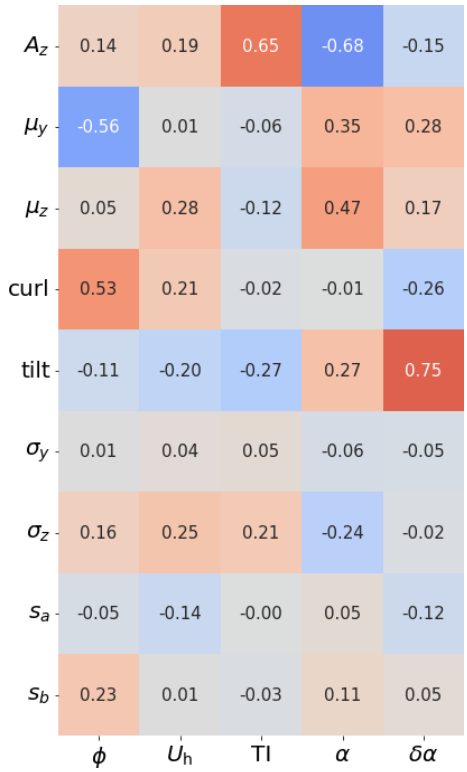

Figure 8. Correlation matrix of the input variables and wake characteristics. A_z is the amplitude of the wake deficit normalized by U_h, μ_y and μ_z the lateral and vertical wake center displacement, "curl" and "tilt" the wake curl and tilt, σ_y and σ_z the width and height of the wake, and s_a and s_b the quadratic and linear wake width parameter.

plays how the nine wake characteristics correlate with the input variables.

The wake center deficit normalized with the hub height wind speed, denoted A_z, is highly correlated with shear α and turbulence intensity TI and shows a moderate correlation with veer $\delta\alpha$, corresponding to the correlations found in previous studies (e.g., Bastankhah and Porté-Agel, 2016; Schottler et al., 2017). A_z has a weak correlation with the hub height wind speed U_h as it has already been used to normalize the deficit.

The lateral wake center displacement μ_y has a relatively high correlation with the yaw misalignment ϕ, confirming that the wake is deflected when the turbine is operated with a yaw misalignment. Moderate correlations with α and $\delta\alpha$ are found, corresponding to previous findings (e.g., Fleming et al., 2015; Sengers et al., 2022) that found that wake deflection is affected by atmospheric conditions. The vertical wake center displacement μ_z, a relatively unexplored wake characteristic, appears to be positively correlated with α. It is hypothesized that this is due to a larger wind speed gradient at lower tip height, increasing the mixing compared to that at upper tip height, effectively moving the wake center upwards. Further analysis (results not shown here) suggests that the vertical wake center displacement, and with that its correlation with input variables, is independent of wind direction. This excludes the influence of topography on these results.

The curl only correlates with ϕ, whereas the wake tilt is highly correlated with $\delta\alpha$, corresponding to Abkar et al. (2018). Lastly, variables related to wake size (σ_y, σ_z, s_a, s_b) have very weak correlations with the input variables, which could be due to the spatial resolution of the lidar PPI scans.

4.1.2 Lateral wake center displacement and wake curl

Two wake characteristics, μ_y and curl, are investigated as a function of ϕ as these are deemed important for wake steering. Figure 9a demonstrates that $|\mu_y|$ typically increases with $|\phi|$; hence the wake deflection is larger for larger yaw misalignment angles, although there is a lot of scatter in the field measurements as also indicated by the correlation coefficient R. Three clusters can be identified, corresponding to the distribution of yaw angles shown in Fig. 4.

To check whether μ_y's order of magnitude is reasonable, field measurements are compared with LES results. Differently than in Sect. 2.3, the turbine simulated here represents turbine T1 in the field, for which the aerodynamic properties were provided by the operator in the Bladed model and translated into FAST. Because of computational restrictions, only three yaw settings $\{-15, 0, 15°\}$ with each of the four inflow conditions were simulated, which will represent only a small part of the full range of conditions observed in the field. The simulations have $U_h \approx 8\,\mathrm{m\,s^{-1}}$, and the inflow variables are $0.11 < \alpha < 0.26$; $1.1° < \delta\alpha < 2.6°$ and $6.0\,\% < \mathrm{TI} < 8.4\,\%$. The LES results show an initial deflection for $\phi = 0°$ (Gebraad et al., 2016), which is not clearly observed in the field. Otherwise, the observed magnitude of deflection is comparable between LES and the field.

Figure 9b displays curl as a function of ϕ. Similarly to μ_y, the field measurements have a larger spread than the LES results, expressed by the lower quality of the linear fit (correlation coefficient R). However, the fitted lines are similar, indicating that the wake curl does indeed occur in the field, something that until now had not conclusively been shown in the literature.

One case is selected (indicated with a white plus sign in Fig. 9) to illustrate what a wake with curl ≈ 0.5 looks like. Figure 10a presents the observed deficit measurements (U_{def}) normalized by U_h, in which the wake's curl is indicated by the dashed black line. The curl is indeed relatively small and could be missed when operating the long-range lidar with a different scanning strategy. Figure 10b represents a reconstructed wake using the composition method (Sect. 3.1), which clearly shows that the wake center has moved to the left and up.

Even though the curling observed in this study is relatively small, Fig. 9b does confirm that the wake curls as expected

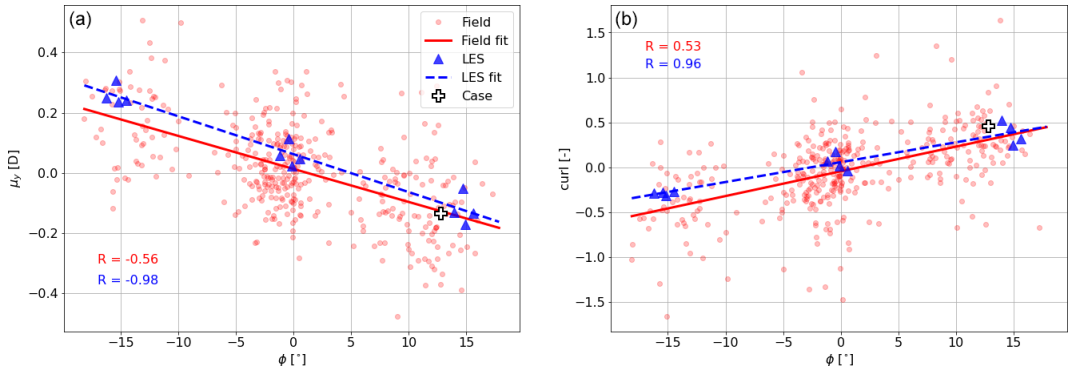

Figure 9. Scatterplot of **(a)** μ_y and **(b)** curl as a function of yaw angle ϕ. Red markers indicate field measurements, and blue triangles indicate LES data. Fitted linear functions are indicated with lines. The quality of these fits is indicated by the correlation coefficient R, corresponding to Fig. 8. White plus signs indicate the case studied in Fig. 10.

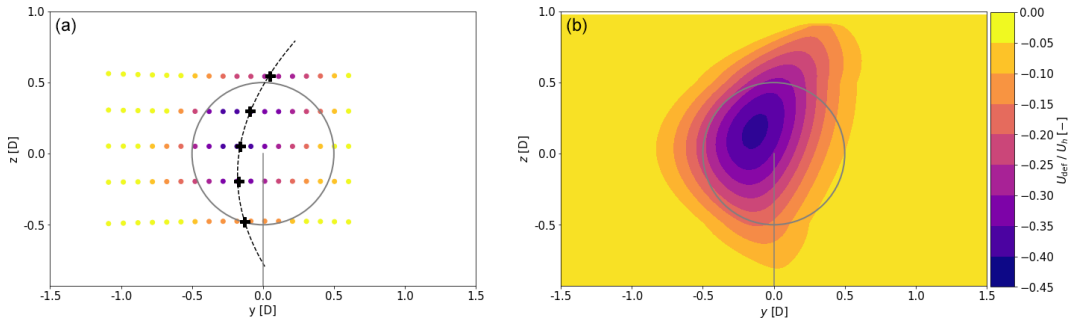

Figure 10. Exemplary case illustrating the wake curl generated by a misaligned turbine ($\phi = 12.8°$, $U_h = 9.4\,\mathrm{m\,s^{-1}}$, $\alpha = 0.33$, $\delta\alpha = -0.8°$, $\mathrm{TI} = 10.4\,\%$). **(a)** The wake deficit of the 10 min averaged lidar data of 17 consecutive PPI scans (colors) and local wake center positions (black plus signs) with corresponding fitted polynomial (dashed black line) indicating the wake curl. **(b)** The wake reconstructed by utilizing the composition method. The color bar applies to both figures.

from numerical and wind tunnel experiments. Field experiments are often restricted to yaw misalignments smaller than 20°, whereas numerical and wind tunnel studies allow for larger misalignments. As suggested in Brugger et al. (2020), this is the reason for the lack of observations of fully curled (or kidney-shaped) wakes in the field.

4.2 Performance of wake models

This subsection presents the performance of the DART and GCH wake models in reproducing the wake characteristics observed in the field. Section 4.2.1 presents how well the models can reproduce the available power measured by the lidar. Following this general result, Sect. 4.2.2 zooms in on how well the models perform under different conditions and Sect. 4.2.3 displays how well the models can reproduce a selection of wake characteristics. Section 4.2.4 discusses how sensitive the models are to the number of training data. Lastly, Sect. 4.2.5 evaluates how well DART performs when only using SCADA data as input.

4.2.1 Comparison of DART and GCH

This subsection discusses the performance of DART and GCH in a comparison with the wake observed in the field. The models were trained (DART, Sect. 3.4.4) or tuned (GCH, Sect. 3.5) on 80 % of the data and are now tested on the remaining 20 % of the data. Figure 11a displays the model accuracy on one resample, using the percentage error (PE) of available power (P_{av}) as a performance metric, analogously to Fig. 7. All models seem to have negative bias (MPE < 0 %), indicating that the rotor-equivalent wind speed U_{eq} is overestimated. DART's bias reduces with an increasing number of input variables, as is evident from the smaller error in DART-7 compared to DART-3 and DART-4. DART-7's bias is comparable to that of GCH. GCH's spread is however larger than DART's, resulting in a larger MAPE despite having a lower MPE.

When repeating this for all 96 resamples, a distribution of MPE and MAPE values can be found (Fig. 11b–c). Also here DART shows a small negative bias ($\overline{\mathrm{MPE}} < 0\,\%$), hence underestimating U_{eq}. GCH has a small positive bias, therefore overestimating U_{eq}, and a much wider distribution.

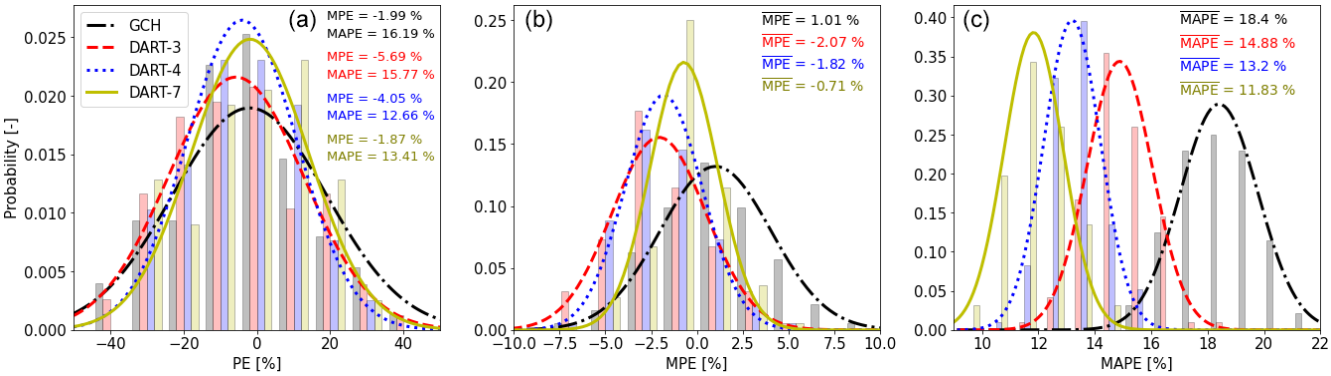

Figure 11. Like Fig. 7 but for the testing data for GCH (black/grey), DART-3 (red), DART-4 (blue) and DART-7 (yellow). Histogram **(a)** for one resample and **(b)** for MPEs and **(c)** for MAPEs of all 96 resamples. Fitted normal distributions are indicated with lines, and statistics are given in the top right.

The distribution of MAPE values indicates that with just three input variables, DART-3 is able to outperform GCH, showing a reduction in $\overline{\text{MAPE}}$ of 19 %. Moreover, both its $\overline{\text{MPE}}$ and $\overline{\text{MAPE}}$ are very similar to those of the training data (Fig. 7), indicating that the model is able to generalize well to unseen or independent data. Adding more variables further improves DART's accuracy, reducing $\overline{\text{MAPE}}$ with 28 % and 36 % for DART-4 and DART-7 compared to GCH. Moreover, the fitted normal distributions of $\overline{\text{MAPE}}$ for GCH and DART-4 or DART-7 hardly overlap, indicating that DART significantly outperforms GCH when trained with at least four variables.

These results show the potential of a data-driven model: more of the variability in wakes observed in the field can be explained by using only four input parameters in a data-driven model than with an industry-standard analytical model.

4.2.2 Model accuracy under different conditions

To gain a better understanding of these results, under what conditions the models' performances differ considerably is investigated. DART-3 is here omitted for brevity. First, the models' errors are investigated in relation to the yaw angle ϕ. Figure 12a displays a histogram of data availability per ϕ bin of 5° over all 96 resamples, while Fig. 12b and c show the MPE and MAPE of P_{av} per bin. The GCH model has MPE > 0 % for $\phi < -7.5°$ and MPE < 0 % for $2.5° < \phi < 7.5°$. This is likely due to low data availability. DART demonstrates a more uniform trend over all yaw angles. When looking at MAPE (Fig. 12c), it can be seen that especially under yawed conditions (both positive and negative), DART seems to outperform GCH. It is hypothesized that this is due to a more accurate estimation of the wake center position in DART.

Figure 12d–f display a similar analysis as a function of shear α. GCH shows an almost linear trend as a function of α, with MPE < 0 % for small α and MPE > 0 % for large α. This indicates that the modeled wake recovery is too slow under low shear and too fast under high shear inflow, which could be due to the turbulence model not explicitly including α as an input parameter. In contrast, DART explicitly uses α to estimate wake characteristics. It therefore produces more uniform results and outperforms GCH, especially when $\alpha > 0.4$ (see Fig. 12c). Over the whole range of ϕ and α, DART-7 is marginally more accurate than DART-4.

4.2.3 Estimating wake characteristics

Finally, the accuracy of GCH and DART in estimating the wake characteristics A_z, μ_y and curl is investigated. The left column of Fig. 13 displays the observed A_z as a function of the model-estimated $\hat{A}_z$. As clearly indicated by the fitted line, the GCH model overestimates small deficits and underestimates large deficits. This could be resolved by giving more weight to outliers when tuning the parameters, although that could lead to the undesirable decrease in accuracy in frequently occurring conditions. The fitted line to the DART-4 results appears to be closer to the unity line, while for DART-7 an even better agreement is found. Additionally, the mean absolute error (MAE) and Pearson correlation coefficient (R) displayed in the top left indicate a more accurate modeling of A_z using DART.

The center column displays a similar analysis for μ_y. The fitted lines imply a higher accuracy for GCH than for DART, while the statistical metrics suggest the opposite. GCH's estimates for μ_y seem to be clustered, which is not true for DART or the measurements. As noted in Sengers et al. (2022), the effect of inflow conditions (e.g., α) on the wake deflection is not well described in GCH. Consequently, the wake deflection is only a function of the yaw angle and the observed clusters can directly be related to the distribution of yaw misalignment angles shown in Fig. 4. Additionally, while in Fig. 13e and h transparent markers can be observed, the markers in Fig. 13b appear to be opaque. Here, many transparent markers overlay each other, indicating that GCH

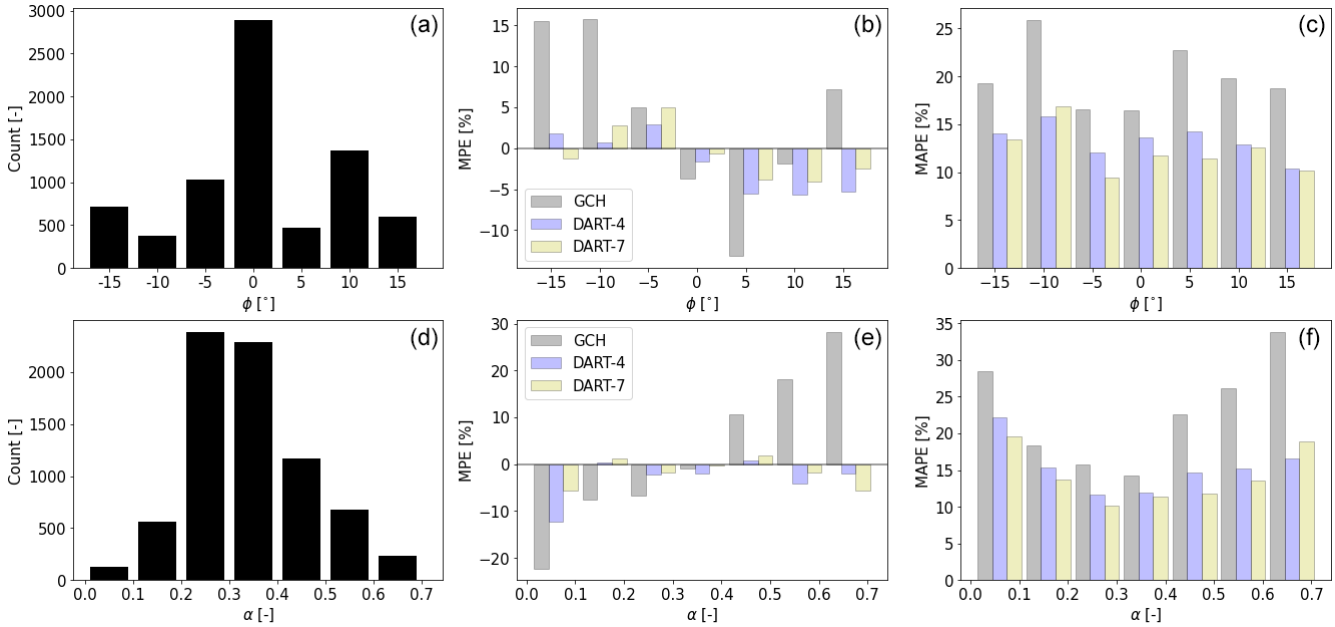

Figure 12. Performance of GCH (grey), DART-4 (blue) and DART-7 (yellow) as a function of ϕ **(a–c)** and α **(d–f)**. Histogram of data availability per bin **(a, d)** and corresponding MPE **(b, e)** and MAPE **(c, f)** per bin.

estimates the same wake center location in all resamples and that these estimates are not affected by the model's tuning parameters.

Lastly, the right column displays the results for curl. GCH does not model any curl, whereas DART is able to capture some of the observed variability. DART-7 performs better than DART-4 as more variables that are (weakly) correlated with curl (see Fig. 8) are considered. Although the variability found in the field is not fully captured by either model, it is clear that the wake curl is better reproduced by DART than by GCH.

4.2.4 Dependency of performance on data set size

An important aspect of data-driven models is understanding how the number of training data affects the model's accuracy. This is especially relevant as one of the most commonly named drawbacks of data-driven models is their high need for data. This subsection studies the sensitivity of the accuracy of DART-4 to the number of training data. DART-7 is not considered due to its long training time. Additionally, GCH is included in the analysis as it contains tuning parameters which could benefit from being tuned to a larger data set. All models are trained with a part of the full data set, ranging from 10 % to 80 %, and tested on the remaining 20 %, analogously to the procedure described in Sect. 3.3. Regardless of the number of training data, the testing data are always 20 % of the original data set and consist of the same cases for fair comparison. When using, e.g., 40 % of the data for training and 20 % for testing, the remaining 40 % are not used at all.

Figure 14 displays the accuracy of the models as a function of the size of the training data set. The metrics used for this analysis are again the distribution of MPE and MAPE values of the 96 resamples, corresponding to the normal distributions shown in Fig. 11. Figure 14a again reveals that DART-4 has a negative bias (MPE < 0 %), which is present regardless of the number of training data, whereas GCH typically has a small positive bias (MPE > 0 %). The uncertainty bands, representing 1 standard deviation indicated by the shaded area, are larger for GCH than for DART-4, while for both models the uncertainty is reduced when trained with more data.

Figure 14b displays the distribution of MAPE values of the 96 resamples as a function of the data set size. GCH and DART-4 have a similar accuracy when few data are available, but DART-4 already outperforms GCH when as little as 20 % of the data set (≈ 75 cases or 13 h) is used for training. Note that this does not indicate 13 h of consecutive measurements but rather 75 cases covering a range of meteorological conditions representative of the variability experienced by the turbine. Additionally, the accuracy of DART-4 seems to continue to improve when adding more data, albeit at a slower rate, whereas GCH hardly shows any improvement with higher data availability.

4.2.5 Performance with SCADA data as input

In this subsection, DART-4 is trained using only data routinely available to the operator (SCADA data) as input variables, called DART-4S. However, it still uses the measurements from the nacelle-mounted lidar to obtain the wake characteristics. Input variables includes power P, rotor speed

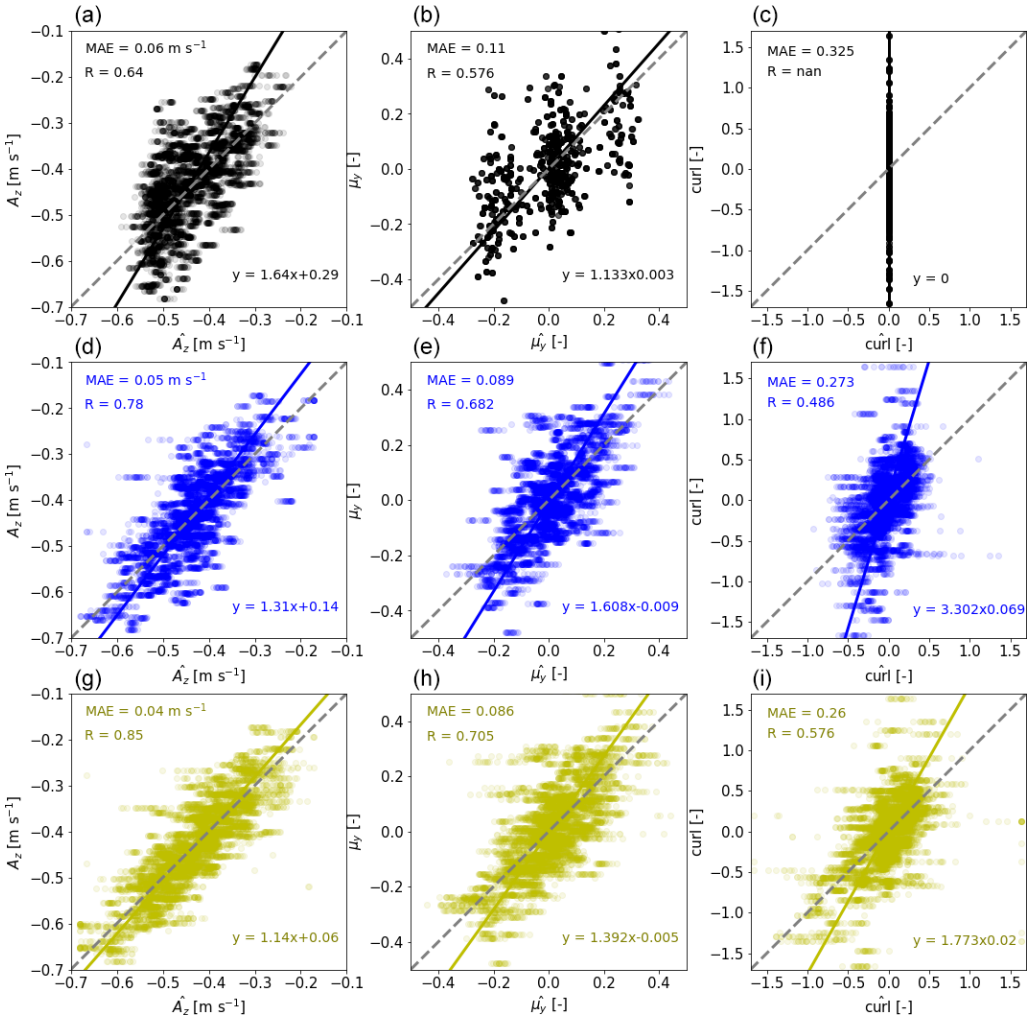

Figure 13. Accuracy of GCH **(a–c)**, DART-4 **(d–f)** and DART-7 **(g–i)** in estimating wake characteristics A_z **(a, d, f)**, μ_y **(b, e, h)** and curl **(c, f, i)**. The models' estimates are given on the x axis and the observations on the y axis. Solid lines indicate linear orthogonal distance regression fits and dashed lines the identity lines.

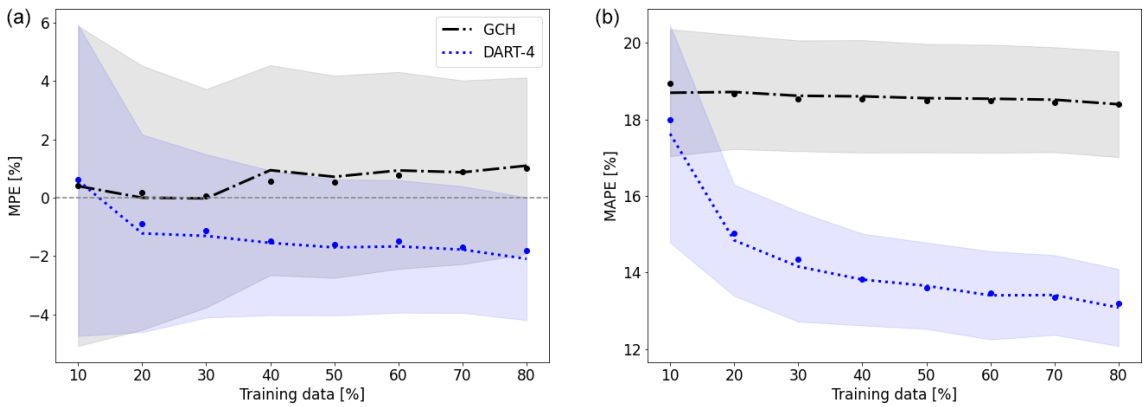

Figure 14. MPE **(a)** and MAPE **(b)** as a function of training data set size. Markers indicate the means ($\overline{\text{MPE}}$ and $\overline{\text{MAPE}}$), lines indicate the median and shaded areas indicate the standard deviation, corresponding to the fitted normal distributions in Fig. 11.

ω, wind speed U_S and turbulence intensity TI_S estimated from the cup anemometer and yaw misalignment ϕ_S extracted from the wind vane. As discussed, according to Fig. 4, this signal contains a systematic bias (Fleming et al., 2021) and is disturbed by the turbine yaw misalignment (Kragh and Fleming, 2012; Simley et al., 2021a), resulting in misalignments being overestimated by the nacelle vane. No correction was applied here, as data-driven models can compensate for any systematic biases. A similar reasoning can be applied to the use of TI_S: although turbulence intensity estimates from a nacelle cup anemometer are affected by the rotor (e.g., Barthelmie et al., 2007), biases can be handled by data-driven models and are therefore acceptable. Additionally, data-driven models are better able to deal with noise than analytical models, which assume that the inflow information is undisturbed or need error terms inserted in the model equations (Schreiber et al., 2020).

The main issue of only using SCADA data is that there is no reliable estimate for the vertical wind speed profile. In this study, the shear α measured at the met mast is estimated from TI_S using the fitted linear relation: $\hat{\alpha} = 0.625 - 0.023\ TI_S$. Because α and TI_S are quite weakly correlated ($R = -0.47$), this simple approach introduces uncertainty. However, developing a more sophisticated solution was deemed out of the scope of this work and this approach is deemed sufficient for the current purpose. An alternative approach could be to use strain measurements from the turbine's blades to estimate shear, as demonstrated in Bertelè et al. (2017, 2021), although this would also involve additional sensors.

Figure 15 displays the results of DART-4S, using $\{\phi_S, TI_S, U_S, P\}$ as input, in a comparison with GCH and DART-4, both trained with met mast data. The accuracy of DART-4S is very similar to DART-4, showing a larger negative $\overline{\text{MPE}}$ but an almost identical $\overline{\text{MAPE}}$. This indicates that using an arguably lower quality data set hardly affects the accuracy of the wake estimates. It is hypothesized that this is because the SCADA data better capture the atmospheric conditions at the turbine, whereas met mast data are subject to heterogeneity between met mast and turbine. This would counter the lower quality of the data, leading to only a slight decrease in the model accuracy. Interestingly, DART-4S is significantly more accurate than GCH, even though the latter needs undisturbed measurements as input.

5 Discussion

Section 5.1 discusses the measurement campaign and its accuracy. In Sect. 5.2 the limitations of the data-driven model are reviewed. Finally, Sect. 5.3 focuses on the implication of this study's results for future work.

5.1 Campaign

This subsection discusses some key takeaways for future campaigns (Sect. 5.1.1) and an uncertainty analysis considering measurement errors (Sect. 5.1.2).

5.1.1 Lessons learned

Several considerations regarding the experimental campaign are noteworthy. First, the nacelle-mounted lidar's scanning strategy was based on Brugger et al. (2019, 2020) and evaluated systematically using large-eddy simulation results and a lidar simulator. However, during this analysis, data losses were not considered. Subsequently, in the field data occasionally all information at one height was filtered out, leaving only information at four heights for the analysis (Sect. 2.3), which could lead to interpolation errors. A more robust approach would have been to perform seven instead of five consecutive PPI scans, although the accuracy of the wake reconstruction method is slightly lower (see Fig. 2). Lastly, in this study only one distance of $4D$ was targeted, but for other purposes it could be desirable to target multiple positions at once. This would likely require more PPI scans with a larger range of elevation angles, as used in Brugger et al. (2020). Large elevation angles are needed to capture the wake close to the turbine, whereas small elevation angles capture the wake further downstream.

Further, no systematic hard-target analysis was performed with the nacelle-based lidar. The horizontal offset relative to the turbine's center axis could be estimated from a set of coarser PPI scans, but no vertical offset could be estimated. Although this is not expected to have a large influence on the results presented here, as is also discussed in Sect. 5.1.2, it is recommended to always carry out a hard-target analysis in future measurement campaigns to reduce the uncertainty in the measurements.

5.1.2 Measurement uncertainty

Although the measurement data after filtering have been considered the "ground truth" in this study, a few aspects affecting the data quality should be considered. Homogeneity of the background flow is assumed, as well as a vertical wind profile that can be described with the power law, which is not always satisfied. This refers specifically to the trees in the wind direction sector around $\delta = 350°$ that are assumed not to affect met mast data, as mentioned in Sect. 2.1. Besides, turbines T3 (induction zone) and T4 (wake) are assumed to not affect the wake, although this cannot be ruled out entirely. Lastly, the lidar measurements are inherently subject to probe volume averaging, and a different filtering method than the one described in Sect. 2.3 will retain other information and therefore result in slightly different wake characteristics.

Additional analyses were carried out to investigate the effect of measurement uncertainty in the results presented in Sect. 4, specifically those in Fig. 11. An overview of

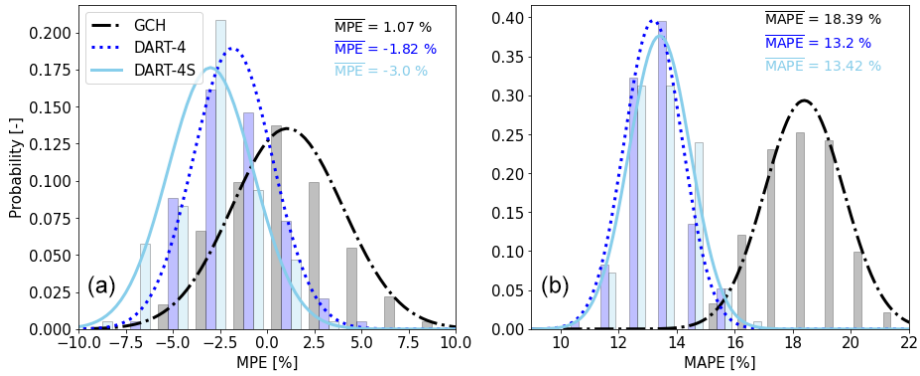

Figure 15. Same as Fig. 11b–c but with DART-4S (light blue).

Table 3. Overview of all tests carried out for the uncertainty analysis.

Test	Description
0	Original; Fig. 11
1	$\phi_{\text{PPI}} - 2.18°$
2	$\phi_{\text{PPI}} - 1.09°$
3	$\phi_{\text{PPI}} + 1.09°$
4	$U_{\text{h}} + 0.2\,\text{m}\,\text{s}^{-1}$ and $U_{\text{lt}} - 0.2\,\text{m}\,\text{s}^{-1}$
5	$U_{\text{h}} - 0.2\,\text{m}\,\text{s}^{-1}$ and $U_{\text{lt}} + 0.2\,\text{m}\,\text{s}^{-1}$
6	$\delta_{\text{h}} + 1.5°$ and $\delta_{\text{lt}} - 1.5°$
7	$\delta_{\text{h}} - 1.5°$ and $\delta_{\text{lt}} + 1.5°$

these tests is displayed in Table 3. First, the influence of the missing hard-target analysis (Sect. 5.1.1) is investigated. In the original measurements, an upward vertical displacement of the wake center of $0.15D$ was observed, averaged over all 382 cases. Although displacements of this magnitude have been observed in numerical simulations (Sengers et al., 2020), it is here assumed that this is purely the consequence of an installation error of the lidar. Such a displacement at $4D$ downstream would come from a downward angle of $\tan^{-1}(0.15/4) = -2.18°$. In test 1, the elevation angles of all lidar scans were adjusted with this value, resulting in an average vertical wake center displacement of zero. Since vertical wake center displacements have been observed in other studies (e.g., Bastankhah and Porté-Agel, 2016; Sengers et al., 2020), the previous test was not deemed completely realistic. Additional tests (2 and 3) with half the correction value, as well as a positive correction value, were carried out. The next set of tests varies the wind speed and direction measured at the met mast. As noted in Sect. 2.5, the accuracy of the anemometer is $0.2\,\text{m}\,\text{s}^{-1}$ and the accuracy of the vane is 1.5°. These values are used in tests 4–7, varying the measurements at hub height (subscript h) and lower tip height (lt) in opposite directions, investigating the maximum influence of these uncertainties. Note that this affects the shear and veer as well.

In all tests, DART-4 (DART-3, DART-7 and DART-4S omitted for brevity) and GCH are trained on the adjusted data for all 96 resembles. Their accuracy on the testing data is evaluated using $\overline{\text{MPE}}$ and $\overline{\text{MAPE}}$ and shown in Fig. 16. Compared to the original (Test 0) of Fig. 11, Test 1 shows a slightly poorer performance of DART-4 (larger $\overline{\text{MPE}}$ and $\overline{\text{MAPE}}$). It is hypothesized that this is due to the fact that now the five PPI scans do not fully target the rotor area anymore, resulting in less relevant information about the wake to estimate the available power. GCH, on the other hand, seems to perform better (smaller $\overline{\text{MPE}}$ and $\overline{\text{MAPE}}$) than the original. Since no vertical displacement is estimated with GCH, these new data more closely resemble the model's assumptions. Tests 2 and 3 confirm this, as GCH's $\overline{\text{MAPE}}$ increases with the magnitude of the vertical wake center displacement.

Tests 4 and 5 illustrate that both models perform better when the shear is decreased (Test 5) compared to when the shear is increased (Test 4), which relates to the fact that more uniform conditions are easier to reproduce. Lastly, GCH performs worse in Test 6 and better in Test 7 compared to the original, but no satisfying explanation was found.

In general, these tests demonstrate that the more closely a data set resembles the model assumptions, the better the model performs. DART-4 typically shows a higher $\overline{\text{MPE}}$ and a lower $\overline{\text{MAPE}}$ compared to GCH, which is similar to the results presented in Sect. 4. This uncertainty analysis is believed to demonstrate that the results presented in this study are robust and not very sensitive to measurement uncertainty.

5.2 Data-driven model

DART's quantitative results presented in this study are not fully generalizable. The fitted coefficients in Eq. (3) are only valid for the scenario considered in this study, and it is unknown how the model's accuracy transfers to different scenarios, such as other turbine types and downstream distances. Besides, while the range of achieved yaw misalignments is typical of field experiments nowadays (e.g., Fleming et al., 2020, 2021; Doekemeijer et al., 2021), future campaigns could see larger misalignments such as those cur-

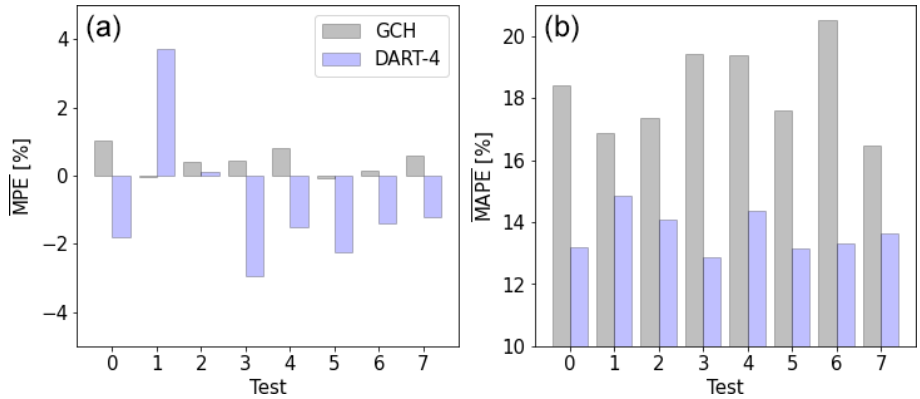

Figure 16. Results of the tests of Table 3. $\overline{\text{MPE}}$ and $\overline{\text{MAPE}}$ of Test 0 correspond to the values found in Fig. 11.

rently considered in numerical studies (e.g., Howland et al., 2016; Martínez-Tossas et al., 2019; Bastankhah et al., 2022). Further lidar measurements would be needed in new scenarios to guarantee accurate model estimates, and although it needs relatively few data for retraining in new situations (Sect. 4.2.4), this limits the potential for application of data-driven models in wake steering control.

To overcome this, future work should attempt to generalize the model's coefficients. Alternatively, model equations using coefficients determined with previous numerical or experimental data could still be used at new locations to generate a first estimate of the wake characteristics. Assuming that the wake position and shape are sufficiently accurately modeled, coefficients for the wake deficit could be retrained using SCADA data by deducing a rotor-equivalent wind speed.

Lastly, other data-driven models could be used. Currently, to the best of our knowledge DART is the only data-driven wake model available that does not make use of complex black-box models such as neural networks. Although it would be interesting to compare different data-driven models, more complex models typically need more data. For instance, Asmuth and Korb (2022) proposed a neural network and showed they need at least 800 cases to train the model for non-yawed cases only. Although their results are extremely promising, extending this to include wake steering would likely require a substantially longer measurement campaign.

5.3 Implications for future work

As noted in the Introduction, the industry appears to be hesitant to adopt the wake steering strategy due to large uncertainties. To overcome this, yaw controllers need to become more sophisticated, for instance by using closed-loop controllers (Doekemeijer et al., 2020; Howland et al., 2020) or using preview information (Simley et al., 2021b). On the other hand, the low-fidelity wake models that are utilized to determine the yaw misalignment set points used by the yaw controller need to become more accurate.

This study contributes to the latter by showing that both DART and GCH perform well on average (small systematic bias) but that DART can capture a higher degree of variability observed in the field. Besides more accurate estimations of the wake deficit, which historically has been the main focus of wake models, this extends to other wake characteristics like wake curl and wake center location. The latter is especially important for wake steering, as erroneous steering can steer the wake into a downstream turbine.

Since DART shows a higher accuracy than GCH in estimating wake characteristics, it can be hypothesized that when using DART to determine yaw misalignment set points used by the yaw controller, the wake steering strategy can be applied more successfully. This can consist of achieving higher power gains when wake steering is performed successfully or reducing power losses due to erroneous steering. However, an extensive campaign would be needed to investigate this, which was considered out of the scope of the current work.

On a more general level, this study shows that data-driven models are a viable alternative to analytical models. Whereas data-driven models have often been criticized for their complex nature, this study has demonstrated that accurate estimations can also be obtained with a very simple linear model.

While the current model focuses on estimated wind speed and consequently power, a similar methodology could be developed to estimate turbulence and consequently turbine loads. Alternatively, it would be interesting to combine analytical and data-driven models in hybrid models. Such models could initially benefit from the robustness of analytical models but exploit the higher accuracy of data-driven models when more data become available.

6 Conclusions

This study uses nacelle-based lidar measurements of the wake of a commercial turbine with a fixed intentional yaw misalignment. Performing a trajectory of five consecutive plan position indicator (PPI) scans with different elevation

angles, a vertical wake cross-section at four rotor diameters downstream is reconstructed. Utilizing the multiple 1D Gaussian method, wake characteristics are obtained. The lateral wake center displacement and wake curl observed in the field compare well with large-eddy simulation results. The results from the lidar measurements demonstrate the occurrence of the wake curl in the field, which had not conclusively been shown in the literature before. This is due to small curling observed for yaw misalignments below 20°, which could be missed when using a different scanning trajectory.

The field measurements are subsequently used to train and validate the DART model and compare the accuracy of the trained data-driven model to the accuracy of the GCH model. When estimating the observed wake characteristics with both wake models, it is demonstrated that DART systematically outperforms GCH. Depending on the number of input variables used for DART, the error is reduced by between 19 % and 36 % compared to GCH. The metric used here is the mean absolute percentage error of the available power of a virtual downstream turbine, averaged over 96 resampled testing data sets. Especially when the turbine is misaligned or high vertical shear is observed, DART outperforms GCH. Besides, DART requires a relatively small number of training data (about 75 cases at specific set points) to outperform GCH. Further analysis suggests that DART's accuracy is hardly affected when only considering SCADA data as input in comparison to using undisturbed measurements from a met mast.

DART shows a high accuracy in the current study, targeting a downstream distance of four rotor diameters and using a range of yaw misalignments commonly used in field experiments. However, these results cannot directly be generalized and further lidar measurements are needed to retrain DART for new scenarios, limiting its applicability. Regardless, this study's results are believed to demonstrate the potential of data-driven wake models and the role they can play in the further deployment of wake steering control strategies.

Appendix A: Lists of abbreviations and symbols

Abbreviations	
APE	Absolute percentage error
CNR	Carrier-to-noise ratio
DART	Data-driven wAke steeRing surrogaTe model
GCH	Gaussian–curl hybrid model
LES	Large-eddy simulation
LOS	Line-of-sight velocity
LUT	Lookup table
MAE	Mean absolute error
MAPE	Mean absolute percentage error
$\overline{\text{MAPE}}$	Mean of the mean absolute percentage errors of all resamples
MM	Meteorological mast
MPE	Mean percentage error
$\overline{\text{MPE}}$	Mean of the mean percentage errors of all resamples
PALM	Parallelized Large-Eddy Simulation Model
PE	Percentage error
PPI	Plan position indicator
R	Pearson correlation coefficient
SCADA	Standard supervisory control and data acquisition
VAD	Velocity–azimuth display
Symbols	
α	Shear
α_{GCH}	Tuning parameter GCH
β_{GCH}	Tuning parameter GCH
δ	Wind direction
δ_{h}	Wind direction at hub height, measured at MM
δ_{lt}	Wind direction at lower tip height
δ_{S}	Wind direction at hub height, measured at nacelle (SCADA)
$\delta\alpha$	Veer
μ_y	Lateral wake center displacement
μ_z	Vertical wake center displacement
ω	Rotor speed
ω_{lidar}	Angular speed of lidar scan
ϕ	Yaw misalignment angle
ϕ_{PPI}	Elevation angle PPI scans
ϕ_{S}	Yaw misalignment angle, measured at nacelle (SCADA)
ϕ_{t}	Target yaw misalignment angle
ϕ_{VAD}	Elevation angle VAD lidar
σ_y	Width of wake at center height
σ_z	Vertical extent of wake
θ_{VAD}	Azimuth angle VAD lidar
A_z	Amplitude of wake deficit normalized with U_{h}
C_{T}	Thrust coefficient
curl	Wake curl
D	Rotor diameter
$k_{a,\text{GCH}}$	Tuning parameter GCH
$k_{b,\text{GCH}}$	Tuning parameter GCH
P	Power
P_{av}	Available power
s_a	Quadratic wake width parameter
s_b	Linear wake width parameter
TI	Turbulence intensity
tilt	Wake tilt
TI_{S}	Turbulence intensity, measured at nacelle
U	Wind speed
U_{def}	Wind speed deficit
U_{eq}	Rotor-equivalent wind speed
U_{h}	Wind speed at hub height; when indicating measurement data, it is measured at MM
U_{lt}	Wind speed at lower tip hub height
U_{S}	Wind speed at hub height, measured at nacelle (SCADA)

Code and data availability. The Data-driven wAke steeRing surrogaTe model (DART), including a short tutorial, is available for download at https://github.com/LuukSengers/DART (last access: 5 May 2023; https://doi.org/10.5281/zenodo.7900032, Sengers and Zech, 2022). The large-eddy simulation results cannot be shared due to confidentiality of the turbine-specific aerodynamic characteristics. A selection of measurement data are available at https://doi.org/10.5281/zenodo.7741395 (Sengers, 2023). This data subset provides 10 min averaged input parameters and lidar scans of cases with $U_h \approx 8\,\mathrm{m\,s^{-1}}$.

Author contributions. BAMS designed the experiment, processed the data, generated the results, and wrote and edited the manuscript. GS provided intensive consultation on the experimental design and generation of the results. PH was heavily involved in organizing the campaign and contributed to the processing of the raw data, as well as translated the aerodynamic properties of the turbine provided by the operator in Bladed into FAST. MK provided general consultation and had a supervisory function. All co-authors reviewed the manuscript.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. The authors would like express their gratitude to eno energy GmbH, specifically to Alexander Gerds, for the opportunity to carry out these experiment on one of their commercial turbines. Carlo Sucameli and Vlaho Petrovic are acknowledged for the implementation of the controller, and Stephan Stone, Anantha Padmanabhan Kidambi Sekar and Jörge Schneemann are thanked for helping during the installation of the sensors. Hauke Beck is recognized for the discussions on the filtering of lidar data. Matthias Zech is thanked for the discussions on the results obtained with DART, and Marijn Floris van Dooren is acknowledged for his help with LiXim. Paul van der Laan and Andreas Rott are thanked for general discussions.

Financial support. The presented work has been carried out within the national research project "CompactWind II" (ref. no. FKZ 0325492H), funded by the Federal Ministry for Economic Affairs and Climate Action (BMWK) on the basis of a decision by the German Bundestag. Computer resources have been provided by the national research project "Heterogener Hochleistungsrechner für windenergierelevante Meteorologie- und Strömungsberechnungen (WIMS-Cluster)" (ref. no. FKZ 0324005), funded by the Federal Ministry for Economic Affairs and Climate Action (BMWK) on the basis of a decision by the German Bundestag.

Review statement. This paper was edited by Jan-Willem van Wingerden and reviewed by three anonymous referees.

References

Abkar, M., Sørensen, J. N., and Porté-Agel, F.: An Analytical Model for the Effect of Vertical Wind Veer on Wind Turbine Wakes, Energies, 11, 1838, https://doi.org/10.3390/en11071838, 2018.

Ahmad, T., Basit, A., Ahsan, M., Coupiac, O., Girard, N., Kazemtabrizi, B., and Matthews, P. C.: Implementation and Analyses of Yaw Based Coordinated Control of Wind Farms, Energies, 12, 1266, https://doi.org/10.3390/en12071266, 2019.

Ainslie, J. F.: Calculating the flowfield in the wake of wind turbines, J. Wind Eng. Ind. Aerod., 27, 213–224, https://doi.org/10.1016/0167-6105(88)90037-2, 1988.

Archer, C. L. and Vasel-Be-Hagh, A.: Wake steering via yaw control in multi-turbine wind farms: Recommendations based on large-eddy simulation, Sustain. Energy Technol. Assess., 33, 34–43, https://doi.org/10.1016/j.seta.2019.03.002, 2019.

Asmuth, H. and Korb, H.: WakeNet 0.1 – A Simple Three-dimensional Wake Model Based on Convolutional Neural Networks, J. Phys.-Conf. Ser., 2265, 022066, https://doi.org/10.1088/1742-6596/2265/2/022066, 2022.

Barthelmie, R. J., Frandsen, S. T., Nielsen, M. N., Pryor, S. C., Rethore, P.-E., and Jørgensen, H. E.: Modelling and measurements of power losses and turbulence intensity in wind turbine wakes at Middelgrunden offshore wind farm, Wind Energy, 10, 517–528, https://doi.org/10.1002/we.238, 2007.

Bastankhah, M. and Porté-Agel, F.: A new analytical model for wind-turbine wakes, Renew. Energ., 70, 116–123, https://doi.org/10.1016/j.renene.2014.01.002, 2014.

Bastankhah, M. and Porté-Agel, F.: Experimental and theoretical study of wind turbine wakes in yawed conditions, J. Fluid Mech., 806, 506–541, https://doi.org/10.1017/jfm.2016.595, 2016.

Bastankhah, M., Shapiro, C. R., Shamsoddin, S., Gayme, D. F., and Meneveau, C.: A vortex sheet based analytical model of the curled wake behind yawed wind turbines, J. Fluid Mech., 933, A2, https://doi.org/10.1017/jfm.2021.1010, 2022.

Bay, C. J., Fleming, P., Doekemeijer, B., King, J., Churchfield, M., and Mudafort, R.: Addressing deep array effects and impacts to wake steering with the cumulative-curl wake model, Wind Energ. Sci., 8, 401–419, https://doi.org/10.5194/wes-8-401-2023, 2023.

Beck, H. and Kühn, M.: Dynamic Data Filtering of Long-Range Doppler LiDAR Wind Speed Measurements, Remote Sens., 9, 561, https://doi.org/10.3390/rs9060561, 2017.

Beck, H. and Kühn, M.: Temporal Up-Sampling of Planar Long-Range Doppler LiDAR Wind Speed Measurements Using Space-Time Conversion, Remote Sens., 11, 867, https://doi.org/10.3390/rs11070867, 2019.

Bertelè, M., Bottasso, C. L., Cacciola, S., Daher Adegas, F., and Delport, S.: Wind inflow observation from load harmonics, Wind Energ. Sci., 2, 615–640, https://doi.org/10.5194/wes-2-615-2017, 2017.

Bertelè, M., Bottasso, C. L., and Schreiber, J.: Wind inflow observation from load harmonics: initial steps towards a field validation, Wind Energ. Sci., 6, 759–775, https://doi.org/10.5194/wes-6-759-2021, 2021.

Blondel, F. and Cathelain, M.: An alternative form of the super-Gaussian wind turbine wake model, Wind Energ. Sci., 5, 1225–1236, https://doi.org/10.5194/wes-5-1225-2020, 2020.

Boccolini, M., Bossanyi, E., Bourne, S., Dombrowski, A., Ferraro, G., Harman, K., Harrison, M., Hille, N., Landberg, L., Levick, T., Manjock, A., Mercer, T., Neubert, A., Ruisi, R., and Skeen, N.: Wind Farm Control: The Route to Bankability, Tech. rep., DNV, https://www.dnv.com/Publications/wind-farm-control-198162 (last access: 10 May 2021), 2021.

Bromm, M., Rott, A., Beck, H., Vollmer, L., Steinfeld, G., and Kühn, M.: Field investigation on the influence of yaw misalignment on the propagation of wind turbine wakes, Wind Energy, 21, 1011–1028, https://doi.org/10.1002/we.2210, 2018.

Brugger, P., Carbajo Fuertes, F., Vahidzadeh, M., Markfort, C. D., and Porté-Agel, F.: Characterization of Wind Turbine Wakes with Nacelle-Mounted Doppler LiDARs and Model Validation in the Presence of Wind Veer, Remote Sens., 11, 2247, https://doi.org/10.3390/rs11192247, 2019.

Brugger, P., Debnath, M., Scholbrock, A., Fleming, P., Moriarty, P., Simley, E., Jager, D., Roadman, J., Murphy, M., Zong, H., and Porté-Agel, F.: Lidar measurements of yawed-wind-turbine wakes: characterization and validation of analytical models, Wind Energ. Sci., 5, 1253–1272, https://doi.org/10.5194/wes-5-1253-2020, 2020.

Carbajo Fuertes, F. and Porté-Agel, F.: Using a Virtual Lidar Approach to Assess the Accuracy of the Volumetric Reconstruction of a Wind Turbine Wake, Remote Sens., 10, 721, https://doi.org/10.3390/rs10050721, 2018.

Doekemeijer, B. M., van der Hoek, D., and van Wingerden, J.-W.: Closed-loop model-based wind farm control using FLORIS under time-varying inflow conditions, Renew. Energ., 156, 719–730, https://doi.org/10.1016/j.renene.2020.04.007, 2020.

Doekemeijer, B. M., Kern, S., Maturu, S., Kanev, S., Salbert, B., Schreiber, J., Campagnolo, F., Bottasso, C. L., Schuler, S., Wilts, F., Neumann, T., Potenza, G., Calabretta, F., Fioretti, F., and van Wingerden, J.-W.: Field experiment for open-loop yaw-based wake steering at a commercial onshore wind farm in Italy, Wind Energ. Sci., 6, 159–176, https://doi.org/10.5194/wes-6-159-2021, 2021.

Fleming, P., Gebraad, P. M. O., Lee, S., van Wingerden, J., Johnson, K., Churchfield, M., Michalakes, J., Spalart, P., and Moriarty, P.: Simulation comparison of wake mitigation control strategies for a two-turbine case, Wind Energy, 18, 2135–2143, https://doi.org/10.1002/we.1810, 2015.

Fleming, P., Annoni, J., Scholbrock, A., Quon, E., Dana, S., Schreck, S., Raach, S., Haizmann, F., and Schlipf, D.: Full-Scale Field Test of Wake Steering, J. Phys.-Conf. Ser., 854, 012013, https://doi.org/10.1088/1742-6596/854/1/012013, 2017a.

Fleming, P., Annoni, J., Shah, J. J., Wang, L., Ananthan, S., Zhang, Z., Hutchings, K., Wang, P., Chen, W., and Chen, L.: Field test of wake steering at an offshore wind farm, Wind Energ. Sci., 2, 229–239, https://doi.org/10.5194/wes-2-229-2017, 2017b.

Fleming, P., King, J., Dykes, K., Simley, E., Roadman, J., Scholbrock, A., Murphy, P., Lundquist, J. K., Moriarty, P., Fleming, K., van Dam, J., Bay, C., Mudafort, R., Lopez, H., Skopek, J., Scott, M., Ryan, B., Guernsey, C., and Brake, D.: Initial results from a field campaign of wake steering applied at a commercial wind farm – Part 1, Wind Energ. Sci., 4, 273–285, https://doi.org/10.5194/wes-4-273-2019, 2019.

Fleming, P., King, J., Simley, E., Roadman, J., Scholbrock, A., Murphy, P., Lundquist, J. K., Moriarty, P., Fleming, K., van Dam, J., Bay, C., Mudafort, R., Jager, D., Skopek, J., Scott, M., Ryan, B., Guernsey, C., and Brake, D.: Continued results from a field campaign of wake steering applied at a commercial wind farm – Part 2, Wind Energ. Sci., 5, 945–958, https://doi.org/10.5194/wes-5-945-2020, 2020.

Fleming, P., Sinner, M., Young, T., Lannic, M., King, J., Simley, E., and Doekemeijer, B.: Experimental results of wake steering using fixed angles, Wind Energ. Sci., 6, 1521–1531, https://doi.org/10.5194/wes-6-1521-2021, 2021.

Fukunaga, K. and Hostetler, L.: The estimation of the gradient of a density function, with applications in pattern recognition, IEEE T. Inform. Theory, 21, 32–40, https://doi.org/10.1109/TIT.1975.1055330, 1975.

Gebraad, P., Teeuwisse, F., Wingerden, J. v., Fleming, P., Ruben, S., Marden, J., and Pao, L.: Wind plant power optimization through yaw controlusing a parametric model for wake effects – a CFDsimulation study, Wind Energy, 19, 95–114, https://doi.org/10.1002/we.1822, 2016.

Herges, T., Maniaci, D., Naughton, B., Mikkelsen, T., and Sjöholm, M.: High resolution wind turbine wake measurements with a scanning lidar, J. Phys.-Conf. Ser., 854, 012021, https://doi.org/10.1088/1742-6596/854/1/012021, 2017.

Howland, M. F., Bossuyt, J., Martínez-Tossas, L. A., Meyers, J., and Meneveau, C.: Wake structure in actuator disk models of wind turbines in yaw under uniform inflow conditions, J. Renew. Sustain. Ener., 8, 043301, https://doi.org/10.1063/1.4955091, 2016.

Howland, M. F., Lele, S. K., and Dabiri, J. O.: Wind farm power optimization through wake steering, P. Natl. Acad. Sci. USA, 116, 14495–14500, https://doi.org/10.1073/pnas.1903680116, 2019.

Howland, M. F., Ghate, A. S., Lele, S. K., and Dabiri, J. O.: Optimal closed-loop wake steering – Part 1: Conventionally neutral atmospheric boundary layer conditions, Wind Energ. Sci., 5, 1315–1338, https://doi.org/10.5194/wes-5-1315-2020, 2020.

Hulsman, P., Sucameli, C., Petroviç, V., Rott, A., Gerds, A., and Kühn, M.: Turbine power loss during yaw-misaligned free field tests at different atmospheric conditions, J. Phys.-Conf. Ser., 2265, 032 074, https://doi.org/10.1088/1742-6596/2265/3/032074, 2022a.

Hulsman, P., Wosnik, M., Petrović, V., Hölling, M., and Kühn, M.: Development of a curled wake of a yawed wind turbine under turbulent and sheared inflow, Wind Energ. Sci., 7, 237–257, https://doi.org/10.5194/wes-7-237-2022, 2022b.

Jensen, N. O.: A note on wind generator interaction, Tech. Rep. Risø-M No. 2411, Risø National Laboratory, https://backend.orbit.dtu.dk/ws/portalfiles/portal/55857682/ris_m_2411.pdf (last access: 17 July 2019), 1983.

Jimenez, A., Crespo, A., and Migoya, E.: Application of a LES technique to characterize the wake deflection of a wind turbine in yaw, Wind Energy, 13, 559–572, https://doi.org/10.1002/we.380, 2010.

Jonkman, J. M. and Buhl Jr., M. L.: FAST User's Guide, Tech. Rep. EL-500-38230, National Renewable Energy Laboratory, https://www.nrel.gov/docs/fy06osti/38230.pdf (last access: 10 April 2019), 2005.

Jonkman, J., Butterfield, S., Musial, W., and Scott, G.: Definition of a 5-MW Reference Wind Turbine for Offshore System Devel-

opment, Tech. Rep. TP-500-38060, National Renewable Energy Laboratory, https://doi.org/10.2172/947422, 2009.

Kanev, S.: Dynamic wake steering and its impact on wind farm power production and yaw actuator duty, Renew. Energ., 146, 9–15, https://doi.org/10.1016/j.renene.2019.06.122, 2020.

King, J., Fleming, P., King, R., Martínez-Tossas, L. A., Bay, C. J., Mudafort, R., and Simley, E.: Control-oriented model for secondary effects of wake steering, Wind Energ. Sci., 6, 701–714, https://doi.org/10.5194/wes-6-701-2021, 2021.

Kragh, K. A. and Fleming, P. A.: Rotor Speed Dependent Yaw Control of Wind Turbines Based on Empirical Data, in: 50th AIAA Aerospace Sciences Meeting, 9–12 January, Nashville, Tennessee, USA, https://doi.org/10.2514/6.2012-1018, 2012.

Krüger, S., Steinfeld, G., Kraft, M., and Lukassen, L. J.: Validation of a coupled atmospheric–aeroelastic model system for wind turbine power and load calculations, Wind Energ. Sci., 7, 323–344, https://doi.org/10.5194/wes-7-323-2022, 2022.

Maronga, B., Banzhaf, S., Burmeister, C., Esch, T., Forkel, R., Fröhlich, D., Fuka, V., Gehrke, K. F., Geletič, J., Giersch, S., Gronemeier, T., Groß, G., Heldens, W., Hellsten, A., Hoffmann, F., Inagaki, A., Kadasch, E., Kanani-Sühring, F., Ketelsen, K., Khan, B. A., Knigge, C., Knoop, H., Krč, P., Kurppa, M., Maamari, H., Matzarakis, A., Mauder, M., Pallasch, M., Pavlik, D., Pfafferott, J., Resler, J., Rissmann, S., Russo, E., Salim, M., Schrempf, M., Schwenkel, J., Seckmeyer, G., Schubert, S., Sühring, M., von Tils, R., Vollmer, L., Ward, S., Witha, B., Wurps, H., Zeidler, J., and Raasch, S.: Overview of the PALM model system 6.0, Geosci. Model Dev., 13, 1335–1372, https://doi.org/10.5194/gmd-13-1335-2020, 2020.

Martínez-Tossas, L. A., Annoni, J., Fleming, P. A., and Churchfield, M. J.: The aerodynamics of the curled wake: a simplified model in view of flow control, Wind Energ. Sci., 4, 127–138, https://doi.org/10.5194/wes-4-127-2019, 2019.

Medici, D., Ivanell, S., Dahlberg, J., and Alfredsson, P.: The upstream flow of a wind turbine: blockage effect, Wind Energy, 14, 691–697, https://doi.org/10.1002/we.451, 2011.

Niayifar, A. and Porté-Agel, F.: Analytical modeling of wind farms: A new approach for power prediction, Energies, 9, 741, https://doi.org/10.3390/en9090741, 2016.

NREL: FLORIS, Version 3.0rc4, GitHub [code], https://github.com/NREL/floris (last access: 6 April 2022), 2022.

Purohit, S., Ng, E. Y. K., and Kabir, I. F. S. A.: Evaluation of three potential machine learning algorithms for predicting the velocity and turbulence intensity of a wind turbine wake, Renew. Energ., 184, 405–420, https://doi.org/10.1016/j.renene.2021.11.097, 2022.

Renganathan, S. A., Maulik, R., Letizia, S., and Iungo, G. V.: Data-driven wind turbine wake modeling via probabilistic machine learning, Neural Comput. Appl., 34, 6171–6186, https://doi.org/10.1007/s00521-021-06799-6, 2022.

Rott, A., Doekemeijer, B., Seifert, J. K., van Wingerden, J.-W., and Kühn, M.: Robust active wake control in consideration of wind direction variability and uncertainty, Wind Energ. Sci., 3, 869–882, https://doi.org/10.5194/wes-3-869-2018, 2018.

Schottler, J., Hölling, A., Peinke, J., and Hölling, M.: Brief communication: On the influence of vertical wind shear on the combined power output of two model wind turbines in yaw, Wind Energ. Sci., 2, 439–442, https://doi.org/10.5194/wes-2-439-2017, 2017.

Schreiber, J., Bottasso, C. L., Salbert, B., and Campagnolo, F.: Improving wind farm flow models by learning from operational data, Wind Energ. Sci., 5, 647–673, https://doi.org/10.5194/wes-5-647-2020, 2020.

Sengers, B. A. M.: Validation of an interpretable data-driven wake model using lidar measurements from a field wake steering experiment, Zenodo [data set], https://doi.org/10.5281/zenodo.7741395, 2023.

Sengers, B. A. M. and Zech, M.: Model code Data-driven wAke steeRing surrogaTe model (DART), Zenodo [code], https://doi.org/10.5281/zenodo.7900032, 2022.

Sengers, B. A. M., Steinfeld, G., Heinemann, D., and Kühn, M.: A new method to characterize the curled wake shape under yaw misalignment, J. Phys.-Conf. Ser., 1618, 062050, https://doi.org/10.1088/1742-6596/1618/6/062050, 2020.

Sengers, B. A. M., Zech, M., Jacobs, P., Steinfeld, G., and Kühn, M.: A physically interpretable data-driven surrogate model for wake steering, Wind Energ. Sci., 7, 1455–1470, https://doi.org/10.5194/wes-7-1455-2022, 2022.

Simley, E., Fleming, P., and King, J.: Design and analysis of a wake steering controller with wind direction variability, Wind Energ. Sci., 5, 451–468, https://doi.org/10.5194/wes-5-451-2020, 2020.

Simley, E., Fleming, P., Girard, N., Alloin, L., Godefroy, E., and Duc, T.: Results from a wake-steering experiment at a commercial wind plant: investigating the wind speed dependence of wake-steering performance, Wind Energ. Sci., 6, 1427–1453, https://doi.org/10.5194/wes-6-1427-2021, 2021a.

Simley, E., Fleming, P., King, J., and Sinner, M.: Wake Steering Wind Farm Control with Preview Wind Direction Information, in: P. Amer. Contr. Conf., 25–28 May, New Orleans, Louisiana, USA, 1783–1789, https://doi.org/10.23919/ACC50511.2021.9483008, 2021b.

Ti, Z., Deng, X. W., and Yang, H.: Wake modeling of wind turbines using machine learning, Appl. Energ., 257, 114025, https://doi.org/10.1016/j.apenergy.2019.114025, 2020.

Tibshirani, R.: Regression Shrinkage and Selection Via the Lasso, J. Roy. Stat. Soc. B. Met., 58, 267–288, https://doi.org/10.1111/j.2517-6161.1996.tb02080.x, 1996.

topographic-map.com: Topographic map, http://topographic-map.com (last access: 1 August 2022), 2022.

Trabucchi, D.: Lidar Measurements and Engineering Modelling of Wind Turbine Wakes, Dissertation, University of Oldenburg, http://oops.uni-oldenburg.de/4671/1/PhD_Thesis_Trabucchi_2019_to_print_v2.pdf (last access: 15 January 2021), 2019.

Trujillo, J. J., Seifert, J. K., Würth, I., Schlipf, D., and Kühn, M.: Full-field assessment of wind turbine near-wake deviation in relation to yaw misalignment, Wind Energ. Sci., 1, 41–53, https://doi.org/10.5194/wes-1-41-2016, 2016.

van Wingerden, J.-W., Fleming, P. A., Göçmen, T., Eguinoa, I., M, D. B., Dykes, K., Lawson, M., Simley, E., King, J., and Astrain, D.: Expert Elicitation on Wind Farm Control, J. Phys.-Conf. Ser., 1618, 022025, https://doi.org/10.1088/1742-6596/1618/2/022025, 2020.

Vollmer, L., Steinfeld, G., Heinemann, D., and Kühn, M.: Estimating the wake deflection downstream of a wind turbine in different atmospheric stabilities: an LES study, Wind Energ. Sci., 1, 129–141, https://doi.org/10.5194/wes-1-129-2016, 2016.

Wagenaar, J., Machielse, L., and Schepers, J.: Controlling Wind in ECN's Scaled Wind Farm, in: Proceeding of

the EWEA Annual Meeting, 16–19 April, Copenhagen, Denmark, https://repository.tno.nl/islandora/object/uuid:7121037f-28df-4aee-8575-89ca278d34eb (last access: 10 July 2021), 2012.

Wang, W., Yang, S., and Yang, Y.: An Improved Data-Efficiency Algorithm Based on Combining Isolation Forest and Mean Shift for Anomaly Data Filtering in Wind Power Curve, Energies, 15, 4918, https://doi.org/10.3390/en15134918, 2022.