Enhancing minute-scale lidar-based power forecasts of offshore wind farms towards an operational use

Frauke Theuer[1,2], Janna Kristina Seifert[1,2], Jörge Schneemann[1,2], and Martin Kühn[1,2]

[1]Carl von Ossietzky Universität Oldenburg, School of Mathematics and Science, Institute of Physics
[2]ForWind - Center for Wind Energy Research, Küpkersweg 70, 26129 Oldenburg, Germany

Correspondence: Frauke Theuer (frauke.theuer@uni-oldenburg.de)

Abstract. Minute-scale power forecasts are gaining importance to support the integration of volatile wind power in particular for offshore wind farms with geographical concentration of generation capacity. Lidar-based approaches have proven useful as an alternative to statistical methods, however, their forecast horizon needs to be extended from currently 5 minutes to at least 15 minutes to be useful for end-users. In this work, we utilize data from an extensive offshore measurement campaign and adapt a lidar-based forecasting approach, e.g. to include wind profile and lidar inclination measurements for improved tilt correction and vertical extrapolation of wind speed, to forecast wind speed and power with horizons of up to 30 minutes. We evaluate individual turbine wind speed and power forecasts and compare them against the benchmark persistence. Further, the impact of forecast characteristics on the forecast skill are analysed. Our results revealed the lidar-based forecast's ability to outperform persistence up to a 16 minute forecast horizon during unstable conditions. An increased wind vector age and propagation duration was found to reduce the forecast skill. Wind farm power forecasts are analysed neglecting large propagation durations, which increased the forecast skill and forecast horizon for which persistence was outperformed. We discuss the applied wind speed extrapolation approaches, the impact of lidar trajectories and wind vector propagation on forecast skill and the value of lidar-based minute-scale power forecast for end-users. In conclusion, the skill of the lidar-based approach highly depends on atmospheric conditions and the forecast characteristics. When considering this for its operational use, the lidar-based forecast has the potential to improve wind farm power forecasts for relevant forecast horizons.

1 Introduction

In order to reach its climate targets the EU aims to cover 45 % of its overall energy consumption by renewable energies by 2030 (European Commission, 2023). In this context, the installed capacity of offshore wind in the EU member states is supposed to increase from 16 GW at present to 111 GW by the end of this decade (WindEurope, 2023). With this increasing share of fluctuating sources that are being fed into our energy system, keeping the grid stable become increasingly challenging. Thus, accurate minute-scale forecasts are required to support the system integration of wind and solar power (Sweeney et al., 2020; Würth et al., 2019). On these time scales, typically statistical methods are used; however, recently remote-sensing-based approaches have proven useful for wind power forecasts (Würth et al., 2019; Theuer et al., 2021, 2022a; Pichault et al., 2021a).

These approaches utilize upstream wind measurements to derive wind speed and power forecasts at a downstream position as
25 a physical-based alternative to statistical methods.

Theuer et al. (2021) developed a lidar-based approach based on a probabilistic radar-based forecast (Valldecabres et al., 2018b, 2020) and applied it to individual free-stream turbines of a large offshore wind farm to predict power with forecast horizons of 5 minutes. The authors further extended this approach to all turbines of the wind farm and by integrating turbine operational data to an observer-based approach (Theuer et al., 2022a). Pichault et al. (2021a) combined a lidar-based forecast

30 with turbine operational data and a machine learning approach to successfully predict the aggregated power of an onshore wind farm.

While these studies have proven the potential of lidar-based forecasting approaches, they are all limited to a forecast horizon of 5 minutes. As Valldecabres et al. (2018b), Würth et al. (2018) and Theuer et al. (2020a) discuss in their work, the short forecast horizon is primarily attributed to limitations in the available data. As follows from the advection methodologies

35 applied, the forecast horizon of lidar-based forecasts is directly related to the measurement range of the lidar devices and the prevailing wind speed. In many cases, the measurement campaigns were not designed for the purpose of minute-scale forecasting, further limiting the forecast horizon.

To our knowledge, the only exceptions with forecast horizons beyond 5 minutes is a study by Würth (2022), in which the authors predicted wind power with forecast horizons of up to 30 minutes, averaged over 10-minute-intervals, at an offshore site

40 and a complex onshore site. At the offshore site, a long-range lidar was positioned on the nacelle of a wind turbine and utilized to predict its power more accurately than persistence for a 10-minute horizon. At the onshore site, the lidar-based approach was neither able to outperform a statistical benchmark for a 0 to 10 minute nor a 10 to 20 minute forecast horizon, which the authors mainly attribute to the complex orography at the site. In order for the forecasts to be useful for applications not only in system integration but also in electricity trading, an extension of the forecast horizon to at least 15 minutes plus an additional

45 processing time of a few minutes for forecast generation and application is required.

The objective of our work is therefore to further increase the forecast horizon and skill of a lidar-based wind power forecast to make it suitable for an operational use. To this end, we conducted an extensive measurement campaign specifically designed for the purpose of lidar-based minute-scale forecasting in an offshore wind farm utilizing a long-range lidar with measurement distances reaching up to 14 km. We extend the forecasting methodology introduced in Theuer et al. (2021) to larger forecast

50 horizons and implement different variants of tilt correction and wind speed extrapolation to make it more robust against missing input data. We evaluate the forecast skill of deterministic wind speed and power forecasts against the benchmark persistence and dependent on forecast horizon and atmospheric conditions. Finally, to reduce the forecast's bias and increase its benefit for end-users we analyse its characteristics in more detail.

2 Methods

55 The work presented in this study is based on the measurement campaign conducted at the offshore wind farm Nordergründe introduced in Section 2.1 that was designed for minute-scale forecasting. The lidar-based forecasting approach developed by
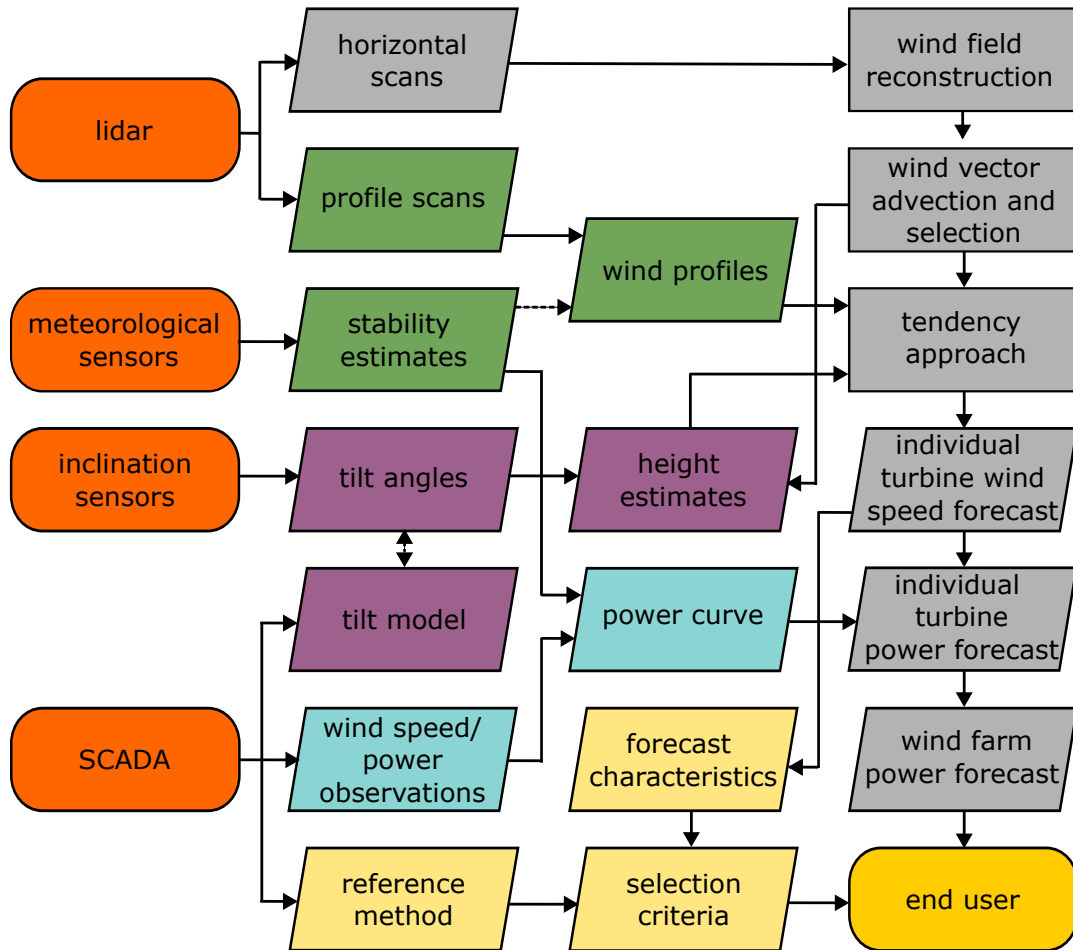
Figure 1. Generic overview of the lidar-based forecasting methodology. The input sensors are shown in orange, the lidar-based method in grey, the wind profile estimation in green, the height estimation in purple, the transformation from wind speed to power forecasts in blue and the forecast assessment and selection in yellow.

Theuer et al. (2021) was applied and extended to larger forecast horizons as explained in Section 2.2. The main steps of the generic forecasting methodology are visualized in Figure 1. Several approaches to measurement height estimation (purple boxes) and wind speed extrapolation from measurement height to hub height (green boxes) are introduced in Section 2.3. The transformation from wind speed to power forecasts is explained in Section 2.4 (blue boxes). Finally, the assessment and selection of the enhanced lidar-based forecasts for an operational use is introduced in Section 2.5 (yellow boxes).

2.1 Measurement campaign at the offshore wind farm Nordergründe

The basis of this work is an extensive measurement campaign at the offshore wind farm Nordergründe in the German North Sea from September 2021 to May 2023. The campaign was specifically designed for minute-scale lidar-based wind speed and

power forecasting and incorporates lessons-learned from previous studies (Theuer et al., 2021, 2022a). Nordergründe consists of 18 turbines of type Senvion 6.2M126 with a rated power of $P_r = 6.15\,\mathrm{MW}$, a hub height of $z_{\mathrm{hh}} = 84\,\mathrm{m}$ and a rotor diameter of $D = 126\,\mathrm{m}$. It is located in close proximity to the coast, with a minimum distance of approximately $15\,\mathrm{km}$ to the German mainland and an average tidal amplitude of approximately $1.4\,\mathrm{m}$ (Figure 2a). A Vaisala Windcube 400S (serial no. 192) was positioned at a height above mean sea level of approximately $z_{\mathrm{TP}} = 16.5\,\mathrm{m}$ on the transition piece (TP) of turbine NG17,

70   indicated by the purple diamond in the wind farm layout visualized in Figure 2b. For the analysis presented in this work, two different scanning trajectories were implemented. First, horizontal PPI scans with an elevation of $0°$, an opening angle of $120°$, an azimuthal resolution of $2°$ and range gates from $300\,\mathrm{m}$ to $13730\,\mathrm{m}$ in $85\,\mathrm{m}$ steps and with a resolution of $150\,\mathrm{m}$ were performed in clockwise direction. The accumulation time was set to $1\,\mathrm{s}$, resulting in a scan duration of $60\,\mathrm{s}$ and an additional measurement reset time of $6\,\mathrm{s}$. Secondly, 15 consecutive PPI scans with elevations of $-0.2°$ and $0°$ to $2.1°$ in $0.15°$ steps were

75   performed with an opening angle of $80°$, an azimuthal resolution of $2°$ and range gates from $300\,\mathrm{m}$ to $9780\,\mathrm{m}$ in $60\,\mathrm{m}$ steps. The accumulation time was set to $0.1\,\mathrm{s}$. The measurement reset time could be reduced to approximately $1\,\mathrm{s}$ by alternating the scanning direction between clockwise and anti-clockwise, resulting in a total scanning duration of approximately $75\,\mathrm{s}$. The purpose of this set of elevated scans is the extraction of wind profiles, which is why we will refer to them as profile-PPIs in the following. More information on the profile-PPIs can be found in Paulsen et al. (2024) and Section 2.3.

80   During the campaign, it was alternated between five horizontal PPIs and a set of profile-PPIs. The scan orientation of the PPI scans was automatically adapted to the five sectors summarized in Table 1. These sectors were optimized for lidar-based forecasting considering the wind direction, the lidar position and the wind farm layout. To avoid a frequent change of trajectories, the wind direction intervals were defined to overlap in a $10°$ sector. In Figure 2 an exemplary scan sector is indicated in blue with the wind direction interval for which this sector was chosen shown in red. Similarly, also the sectors of

85   profile-PPIs were automatically adapted to the wind direction (Table 1). The lidar's view was blocked by the turbine tower for the sector $3°$ to $103°$. For wind directions $5°$ to $115°$, trajectories for a different purpose were measured and thus no lidar-based forecasts generated.

The lidar was equipped with two inclination sensors (Micro-Epsilon INC5701), measuring pitch and roll of the device with a resolution of $2\,\mathrm{Hz}$. Moreover, two meteorological sensors (Vaisala HMP155 and PTB330) measuring air temperature, air

90   pressure and humidity were installed at the lidar location. Two infrared sensors (Heitronics CT09 and CT15) measured the sea surface temperature (SST) with $1\,\mathrm{Hz}$ resolution and applied a correction for sky radiation. As a complement to the SST measurements, OSTIA data was used (Good et al., 2020).

For forecast generation and evaluation $0.2\,\mathrm{Hz}$ SCADA data, including wind speed, wind direction, power, blade pitch, yaw angle and turbine operational status, were available.

95   In Figure 1 the utilized sensors are visualized as orange boxes. Not all sensors and measurements were available for the whole duration of the campaign. Thus, Table 2 provides an overview over all used data and their availability throughout the campaign. It also summarizes what each data set is used for, which will be explained in more detail in the following sections.

Table 1. Lidar trajectories for different wind direction sectors.

wind direction [°]	lidar azimuth [°]	
	horizontal PPI	profile PPI
115 to 185	103 to 223	
115 to 140		103 to 183
140 to 160		110 to 190
160 to 180		130 to 210
175 to 215	140 to 260	
180 to 200		150 to 230
205 to 245	175 to 295	
200 to 220		170 to 250
220 to 240		190 to 270
235 to 275	210 to 330	
240 to 260		210 to 290
265 to 5	243 to 3	
260 to 280		230 to 310
280 to 300		250 to 330
300 to 320		270 to 350
320 to 5		283 to 3

Table 2. Summary of senors utilized for the analysis presented in this study. For each sensor the considered variables, period with available data and application is shown.

sensor	variables	period	application
Windcube400S	horizontal v_{LOS}	09/2021–09/2022, 01/2023–05/2023	wind vector propagation
	elevated v_{LOS}	09/2021–09/2022	wind speed profile measurements
Micro-Epsilon INC5701	pitch, roll	04/2022–05/2023	estimation of lidar tilt
SCADA data	wind speed, power	09/2021–05/2023	
Vaisala PTB330	pressure	09/2021–05/2023	air density correction
Vaisala HMP155	air temperature, humidity	09/2021–05/2023	
Heitronics CT09/CT15	sea surface temperature	04/2022–05/2023	atmospheric stability
OSTIA	sea surface temperature	09/2021–05/2023	
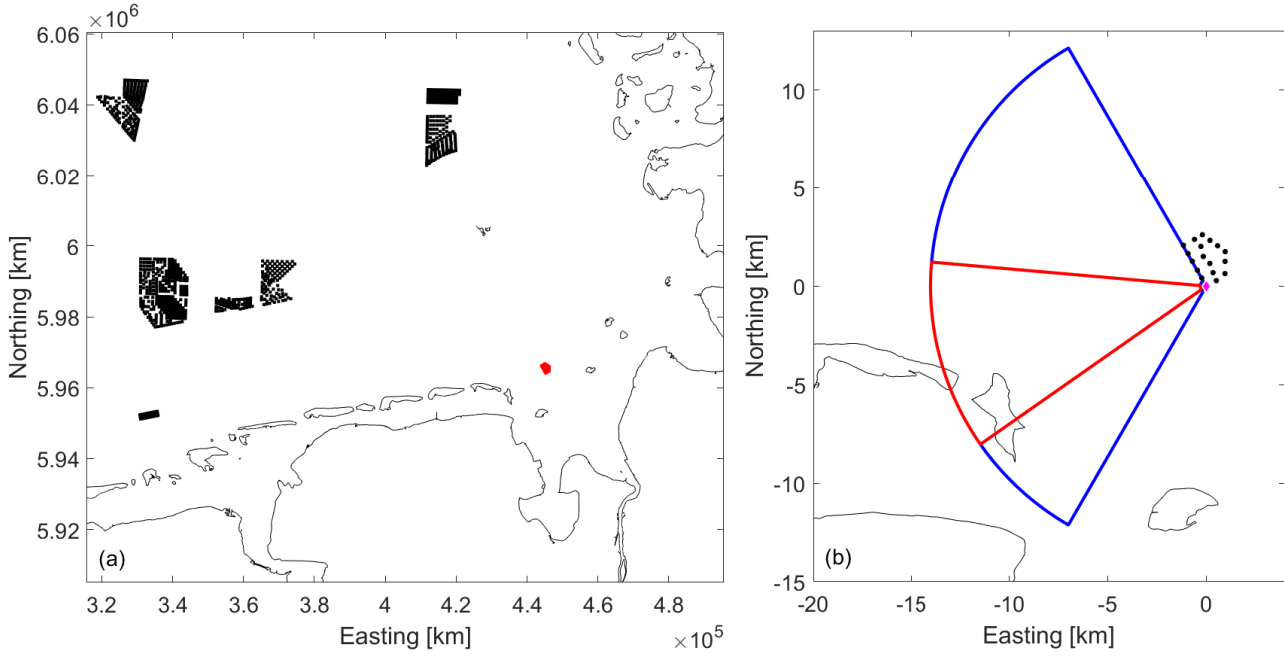
Figure 2. (a) Location of the offshore wind farm Nordergründe in red and other offshore wind farms in the German North Sea (2023) in black. In (b) the layout of Nordergründe is shown with the lidar position as origin of the coordinate system and marked as purple diamond. The azimuth scan sector is depicted in red. Blue lines indicate the corresponding wind direction interval for which this sector was chosen.

2.2 Lidar-based wind speed forecasting

The lidar-based forecast (LF) presented in this work is based on the method developed by Theuer et al. (2021). It is summarized by grey boxes in Figure 1. Horizontal single lidar PPI scans (see Section 2.1) are used to derive wind speed information at hub height at individual turbine positions downstream. To this end, the lidar scans are first filtered using a quality flag provided by the manufacturer of the device. Applying a Velocity Azimuth Display (VAD)-like fit the horizontal wind direction χ dependent on range gate r is then determined for each individual lidar scan (Werner, 2005). This is based on the assumption that the vertical wind speed component can be neglected and the wind direction is homogeneous across each range gate. The measured line-of-sight (LOS) wind speed u_{LOS} is further translated to horizontal wind speed u_{hor} considering χ, the elevation φ and the azimuth angle ϑ of the lidar using

$$u_{\mathrm{hor}}(r,\vartheta) = \frac{u_{\mathrm{LOS}}(r,\vartheta)}{\cos(\vartheta - \chi(r))\,\cos(\varphi)}. \tag{1}$$

Areas of the scan with an almost perpendicular orientation of azimuth angle and wind direction, i.e. $75° < |\vartheta - \chi| < 105°$, are neglected as they are subject to large uncertainty. For the following analysis, only lidar scans with an overall availability, i.e. the share of valid measurement points per lidar scan after filtering, of at least 50 % were considered.

Both horizontal wind speed and direction, referred to as wind vectors in the following, were interpolated to a Cartesian grid with 50 m spacing and then propagated in space and time using Lagrangian advection (Valldecabres et al., 2018a). Hereby, it is assumed that wind vectors travel with their local wind speed and direction and do not change their trajectory while travelling. Wind vectors, in the following denoted by index v, reaching a pre-defined area of influence (AoI) around the target turbine within the time interval $k \pm 30\,\mathrm{s}$, with forecast horizon k, are selected to contribute to the wind speed forecast. In this study, we chose $2\,D = 252\,\mathrm{m}$ as the AoI's radius following previous work (Theuer et al., 2022a). For a forecast to be considered valid, a minimum of 20 wind vectors was required to contribute. It should be noted that the travelling time of wind vectors contributing to the same forecast might differ due to two reasons. One, several lidar scans measured previous to the initialization time of the forecast t, which is defined as the end time of the last available lidar scan, were considered. And two, also the measurement time within one scan differs due to the finite scanning speed and was taken into account during wind vector propagation.

2.3 Wind speed extrapolation to hub height

After wind vector selection, wind speed forecasts need to be extrapolated to hub height. Here, we applied a method introduced as tendency-based approach in previous work (Theuer et al., 2021). Wind vectors are first extrapolated to the average measurement height of a lidar-based forecast initiated at $t - k_{\mathrm{tend}}$, with k_{tend} being the lead time of the forecast used to derive the wind speed tendency. The choice of k_{tend} can be understood as a trade-off between shorter lead times and thus lower forecasting errors and lead times close to the target lead time k_{targ} and thus higher correlations of the respective forecasting errors. In this work, k_{tend} was optimized in a sensitivity study and defined as

$$k_{\mathrm{tend}} = \begin{cases} 1\,\mathrm{min} & k_{\mathrm{targ}} < 14\,\mathrm{min} \\ k_{\mathrm{targ}} - 13\,\mathrm{min} & k_{\mathrm{targ}} \geq 14\,\mathrm{min}. \end{cases} \tag{2}$$

To extrapolate wind vectors to a common height, first the measurement height of each wind vector needs to be determined. We used the inclination sensors to determine the average pitch and roll value for each lidar scan. Considering pitch, roll, the location of origin of the wind vectors and the curvature of the Earth, a measurement height was calculated. In case no data from the inclination sensors were available (see Table 2), a thrust model developed by Rott et al. (2022) was applied as suggested by Theuer et al. (2021). The model uses averaged wind speed and turbine orientation SCADA data to determine pitch and roll angles for each lidar scan. It was calibrated for the lidar measurements at Nordergründe using the inclination sensor measurements. The measurement height estimation is indicated by the purple boxes in Figure 1.

Once the measurement height of each wind vector $z_{\mathrm{m},v}$ is determined, it can be extrapolated using wind speed profiles. Here, we utilized the profile-PPI lidar scans to derive wind speed profiles. To do so, each individual profile-PPI scan was filtered and its horizontal wind speed reconstructed as described in Section 2.2. In this case, one VAD fit was performed for a set of 10 range gates, thus a 600 m range band. After also correcting the measurement height of the profile-PPI scans as described earlier, horizontal wind speeds were binned according to their measurement heights from 5 m to 365 m in 10 m intervals. Horizontal wind speeds were then averaged for each height band and with 30 min resolution (Paulsen et al., 2024). In a next step, wind speed at the measurement and target height were extracted from the respective wind profile by means of linear interpolation.

In the context of the lidar-based forecasts, these correspond to the profile wind speed $u_{\mathrm{prof},v}$ at measurement height of the wind vectors selected for forecasts initiated at $t - k_{\mathrm{tend}}$ and t respectively. That means, wind speed is not extrapolated from measurement height to hub height, but measurement heights of the two forecasts used to determine the wind speed tendency are aligned.

A height correction parameter $c_{\mathrm{height},v}$ is defined as

$$c_{\mathrm{height},v}(t) = \frac{\overline{u}_{\mathrm{prof}}(t - k_{\mathrm{tend}})}{u_{\mathrm{prof},v}(t)} \tag{3}$$

with the average profile wind speed at measurement height $\overline{u}_{\mathrm{prof}}$.

When no profile-PPIs were available, stability-corrected logarithmic wind speed profiles were defined using the Obukhov length L and roughness length z_0. L and z_0 were defined using pressure, humidity and air temperature measurements at TP height and sea surface temperature measurement as described in more detail in Schneemann et al. (2021) and Theuer et al. (2021). If available, the SST measurements were utilized. Otherwise, the interpolated OSTIA SST was used (see Table 2). Following the work of Theuer et al. (2021) and assuming atmospheric conditions do not change significantly from $t - k_{\mathrm{tend}}$ to t, in this case c_{height} yields

$$c_{\mathrm{height},v}(t) = \frac{\ln\left(\frac{\overline{z}_{\mathrm{m}}(t - k_{\mathrm{tend}})}{z_0}\right) - \Psi\left(\frac{\overline{z}_{\mathrm{m}}(t - k_{\mathrm{tend}})}{L}\right)}{\ln\left(\frac{z_{\mathrm{m},v}(t)}{z_0}\right) - \Psi\left(\frac{z_{\mathrm{m},v}(t)}{L}\right)} \tag{4}$$

with the wind vectors measurement height $z_{\mathrm{m},v}$, the average measurement height $\overline{z}_{\mathrm{m}}$ and the stability correction term Ψ. The wind profile estimation is indicated by green boxes in Figure 1.

After adjusting the wind speed forecast initiated at t to the height of that initiated at $t - k_{\mathrm{tend}}$ using the height correction parameter, their wind speed tendency can be derived. The forecasts $u_{\mathrm{fc},v}$ initiated at times $t - k_{\mathrm{tend}}$ and t predict wind speed at times t and $t + k_{\mathrm{targ}}$ and their tendency can thus be applied to the hub height wind speed observed at the initialisation time t of the forecast $u_{\mathrm{obs}}(t)$, i.e. SCADA wind speed, to estimate the wind speed at hub height at the target time $u_{\mathrm{hh},v}(t + k_{\mathrm{targ}})$. It yields

$$u_{\mathrm{hh},v}(t + k_{\mathrm{targ}}) = c_{\mathrm{height},v}(t) \frac{u_{\mathrm{fc},v}(t)}{\overline{u}_{\mathrm{fc}}(t - k_{\mathrm{tend}})} u_{\mathrm{obs}}(t) \tag{5}$$

with the average wind speed forecast $\overline{u}_{\mathrm{fc}}$. Finally, the wind speed forecast at hub height and the target time is defined as the average of all contributing wind vectors

$$u_{\mathrm{hh}}(t + k_{\mathrm{targ}}) = \frac{1}{N_{\mathrm{vec}}(t + k_{\mathrm{targ}})} \sum_v u_{\mathrm{hh},v}(t + k_{\mathrm{targ}}) \tag{6}$$

with the number of contributing vectors N_{vec}. While in this work we analysed deterministic forecasts only, the introduced method can also be applied to derive probabilistic forecasts. In that case, the selected wind vectors are resampled to a predefined number of ensemble members as suggested by Theuer et al. (2022a). To further improve probabilistic forecast skill, additional post-processing can be performed, e.g. using a calibration approach (Thorarinsdottir and Gneiting, 2010).
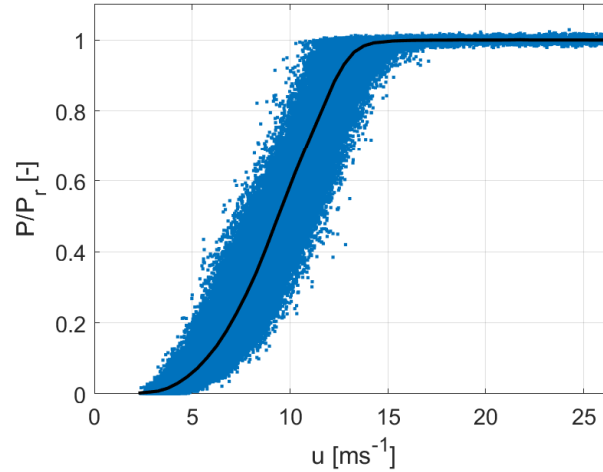
Figure 3. Deterministic air density-corrected power curve of turbine NG5 shown in black. Blue dots visualize 1-minute mean SCADA data used to determine the power curve.

2.4 Wind turbine and wind farm power forecast

In a final step wind speed forecasts for each individual turbine, time step and forecast horizon are transformed to power forecasts, summarized by the blue boxes in Figure 1. Deterministic power curves were defined for each turbine of Nordergründe using 1-minute wind speed and power averages determined from 0.2-Hz-SCADA data. To do so, we excluded situations with behaviour differing from the normal operation, such as curtailment, using pitch data and turbine operational mode. Wind speed values were air density corrected following IEC 61400-12-1 (IEC, 2017) as described in Ulazia et al. (2019). Power values were then averaged within $0.5\,\mathrm{m\,s^{-1}}$ wind speed bins. An exemplary power curve for one turbine is visualized in Figure 3.

To account for the large scatter of the actual power values around the deterministic power curve, power forecasts were then generated in a two-step process. First, utilizing the defined deterministic power curves and linear interpolation, a power value at the target time $P(t+k_{\mathrm{targ}})$ was derived from each air density-corrected wind speed forecast $u_{\mathrm{hh}}(t+k_{\mathrm{targ}})$. Secondly, the same procedure was applied to the hub height SCADA wind speed observed at the initialization time $u_{\mathrm{obs}}(t)$ and yields $P_{\mathrm{pc}}(t)$. A bias correction based on the assumption that the current deviation of the observed SCADA power from the power curve persists was then applied

$$P_{\mathrm{corr}}(t+k_{\mathrm{targ}}) = P(t+k_{\mathrm{targ}}) + (P_{\mathrm{obs}}(t) - P_{\mathrm{pc}}(t)) \tag{7}$$

with $P_{\mathrm{obs}}(t)$ the power observed at the initialization time and $P_{\mathrm{corr}}(t+k_{\mathrm{targ}})$ the bias-corrected power forecast at the target time. Applying this bias correction allows to account for e.g. yaw misalignment or the impact of wind shear or turbulence that might cause deviations from the deterministic power curve, assuming that these effects change on time scales longer than the considered forecast horizon.

To generate a wind farm power forecast we only considered situations during which for at least 12 of the 18 turbines a valid forecast was available. We normalized the aggregated wind power forecasts with the rated power of the aggregated turbines to account for the different number of turbines considered.

2.5 Selection of enhanced lidar-based power forecasts

195 To incorporate the enhanced lidar-based power forecasts into operational forecasting systems, it is important to understand their advantages and shortcomings and performance in comparison to commonly used statistical methods. We therefore analysed the dependency of wind speed forecast skill on the forecast characteristics.

As visualized in Figure 1 by the yellow boxes, the forecast characteristics are derived from individual turbine wind speed forecasts. By evaluation against wind speed observations and other statistical methods, selection criteria can then defined.

200 Hereby, it is important to choose characteristics that can directly be derived from the forecast itself. Other parameters might in principle also be useful for forecast evaluation, e.g. wind speed observations at the target time (see Section 3.2), however, they will only be valuable for forecast selection if available in real-time.

While we evaluated a number of parameters, including, among others, the number of wind vectors contributing to a forecast, the wind vectors' wind speed spread and the number of lidar scans from which wind vectors are selected, in this analysis we 205 focus on the wind vector age as it showed most impact on forecast skill.

We defined the average wind vector age $\overline{\Delta T}$ as the average time difference between a vector's measurement time t_v and the initialisation time t of a forecast:

$$\overline{\Delta T} = \frac{1}{N_{\text{vec}}(t + k_{\text{targ}})} \sum_v (t - t_v) \tag{8}$$

The wind vector age does not consider the propagation duration related to different forecast horizons.

210 Once selection criteria based on the forecast characteristics are defined, they can be applied to select those forecasts that are likely useful to end-users.

3 Results

After introducing the different wind speed extrapolation approaches in Section 2, we evaluate them in comparison to a reference case in Section 3.1. The forecast skill of the LF dependent on atmospheric stability is evaluated for forecast horizons up to 215 30 minutes in Section 3.2. We present different forecast characteristics and their impact on forecast accuracy in Section 3.3. Finally, power forecasts of individual turbines and the aggregated wind farm are evaluated in Section 3.4, also considering the identified selection criteria.

3.1 Evaluation of wind speed extrapolation approaches

In this section we analyse the impact of measurement height correction, i.e. accounting for the lidar tilt and wind speed 220 extrapolation, on the forecast accuracy. The reference case does not account for differences in measurement height between

consecutive wind speed forecasts. Further, three different ways to correct the measurement height are analysed, namely i) considering only the curvature of the Earth, ii) applying the tilt model developed by Rott et al. (2022) and iii) performing a tilt correction utilizing the inclination sensor data. All three cases are evaluated in combination with two extrapolation approaches, that is i) using a stability-corrected logarithmic wind speed profile and ii) utilizing the profile-PPIs. Overall, this amounts to

225   seven different cases. Forecasts are evaluated using the root-mean-squared error (rmse)

$$\mathrm{rmse} = \sqrt{\frac{1}{N_\mathrm{t}} \sum_{i=1}^{N_\mathrm{t}} (\mathrm{fc}_i - \mathrm{obs}_i)^2}. \tag{9}$$

with forecasts fc_i, observations obs_i and the number of available forecasts N_t. Observations are 1-minute-mean wind speed SCADA data.

The six different cases of measurement height correction and wind speed extrapolation are then evaluated against the refer-

230   ence case by means of the rmse skill score (ss)

$$\mathrm{rmse\ ss} = 100\,\% \left(1 - \frac{\mathrm{rmse}}{\mathrm{rmse}_\mathrm{ref}}\right) \tag{10}$$

with the reference case's rmse $\mathrm{rmse}_\mathrm{ref}$. Negative rmse skill scores indicate a lower skill compared to the reference.

For this analysis, only periods with all necessary data available were considered (see Table 2), reducing the available data set to a period of approximately six months from April 2022 to September 2022. Figure 4 shows the rmse ss for the six cases

235   analysed and averaged for all 18 turbines of NG. All of the extrapolation approaches can improve the forecast skill for most of the forecast horizons compared to the reference. The highest rmse ss of up to 2.6 % at 15 minute forecast horizon can be observed when applying the tilt model in combination with the measured profiles. The combination of inclination sensors and measured profiles shows an only slightly lower skill score. For all cases, measured profiles outperform the logarithmic profiles. This effect is in particular distinct for larger forecast horizons to which more measurements at larger ranges and with a larger

240   spread of the measurement height contribute. For forecast horizons up to 14 to 15 minutes the rmse ss increases before it starts to decrease with further increasing lead times. When using logarithmic wind profiles and a tilt correction, skill scores reach negative values from 24 minutes onwards.

3.2   Wind speed forecast skill dependent on atmospheric stability and forecast horizon

After the comparison of the different wind speed extrapolation approaches in Section 3.1, we analyse the whole available data

245   set from 09/21 to 05/23, excluding 09/22 to 01/23 due to non-available horizontal lidar scans (see Table 2), in the following. Despite the slightly better results for the tilt model shown in Section 3.1, we rely on pitch and roll measurements for height estimation if available and use the results of the tilt model as an alternative. This choice is related to the availability and calibration of the model and will be discussed in more detail in Section 4.1. For wind speed extrapolation, wind profile measurements are used as a first option and stability-corrected logarithmic wind profiles as a back-up. If no lidar tilt or no wind speed profile can

250   be determined, no forecast is generated. In total, 92.186 valid wind speed forecasts for individual wind turbines were generated for a forecast horizon of 1 minute in stable and 146.602 in unstable cases. We bin the stratification in only two classes and
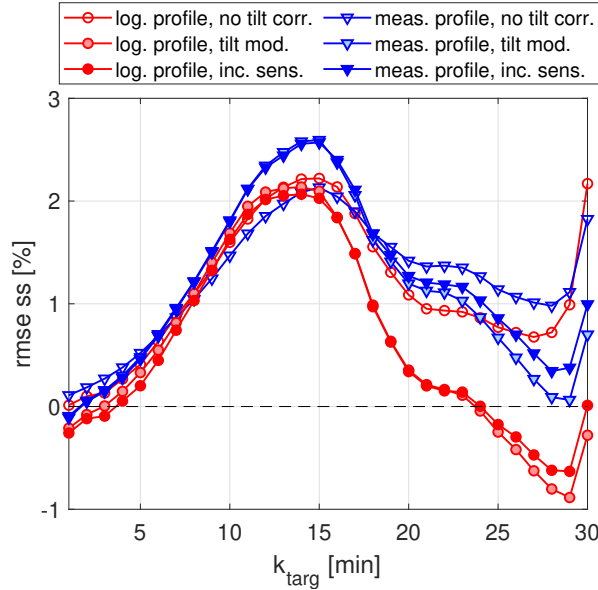
Figure 4. The rmse ss in % for forecast horizons up to 30 minutes and the six different combinations of tilt correction and extrapolation approaches. Forecasts without height correction are used as a reference. The black dashed line indicates a skill score of 0 %.

denote all cases with $L > 0\,\mathrm{m}$ as stable and $L < 0\,\mathrm{m}$ as unstable, recognising that the forecast characteristics for neutral stratification, e.g. classified by $|L| > 500\,\mathrm{m}$, could differ. With increasing lead time, the forecast availability significantly reduces. For a 15-minute forecast horizon the number of valid forecasts yields 60.071 in stable and 114.417 in unstable cases, for a

255   30-minute horizon only 15.572 and 29.804, respectively. Wind speed forecasts are evaluated in terms of rmse and bias

$$\mathrm{bias} = \frac{1}{N_\mathrm{t}} \sum_{i=1}^{N_\mathrm{t}} (\mathrm{fc}_i - \mathrm{obs}_i). \tag{11}$$

Moreover, the LF is compared against the benchmark persistence using the rmse ss with persistence as reference method. Persistence is a commonly used statistical method that assumes the future value equals the current observation.

The wind conditions during the analysed period are visualized in Figure 5. The wind rose shows 1-minute-mean wind speed

260   and direction SCADA data averaged over all turbines of the wind farm for time steps with valid lidar-based forecasts. As described in Section 2.1, wind direction sectors from $5°$ to $115°$ are not covered by the lidar scans and thus not included in the analysis.

In Figure 6a-c we evaluate the wind speed LF's skill against persistence for different forecast horizons and distinguishing between atmospheric conditions. In Figure 6a the average normalized rmse for all turbines of NG of both persistence (empty

265   markers) and the LF (filled markers) are shown in blue for unstable and in red for stable atmospheric conditions. For the LF, additionally, the individual turbines' scores are visualized as pale-coloured lines. The rmse ss and normalized bias are shown
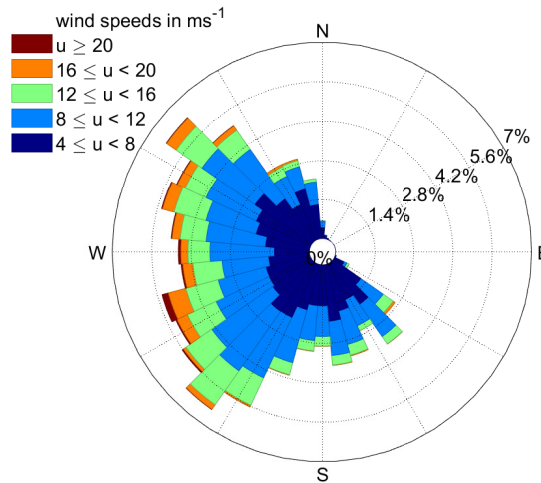
Figure 5. Wind speed and wind direction distribution of all analysed cases. Shown are 1-minute mean SCADA values averaged over all turbines.
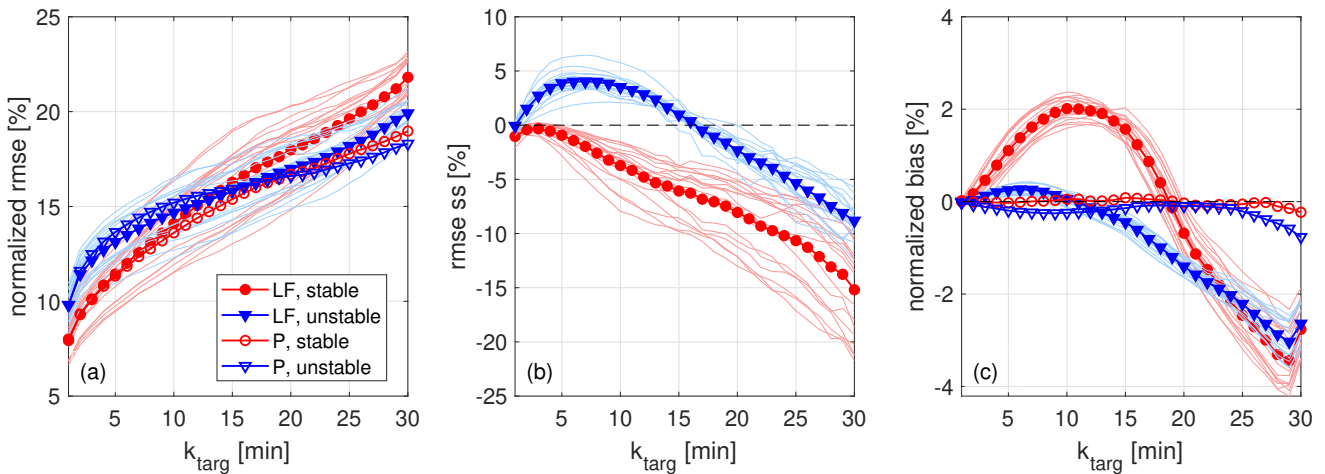
Figure 6. (a) Normalized rmse and (c) bias of the wind speed LF (filled markers) and persistence (P, empty markers) for forecast horizons up to 30 minutes and distinguishing between stable cases as red circles and unstable cases as blue triangles. Thick lines with markers indicate the turbines' average scores while thin lines without markers show scores of individual turbines. In (b) the rmse skill score in % is shown. Black dashed lines indicate a 0 % skill score and normalized bias, respectively.

in Figure 6b and c, respectively. The rmse and bias are normalized with the average wind speed of the analysed time steps that was determined from SCADA data. A negative bias indicates an underestimation of wind speed.

As shown in Figure 6a the rmse increases for both persistence and the LF with increasing forecast horizon. For small lead times, forecasts are more accurate during stable situations as compared to unstable ones. This is in good agreement with results
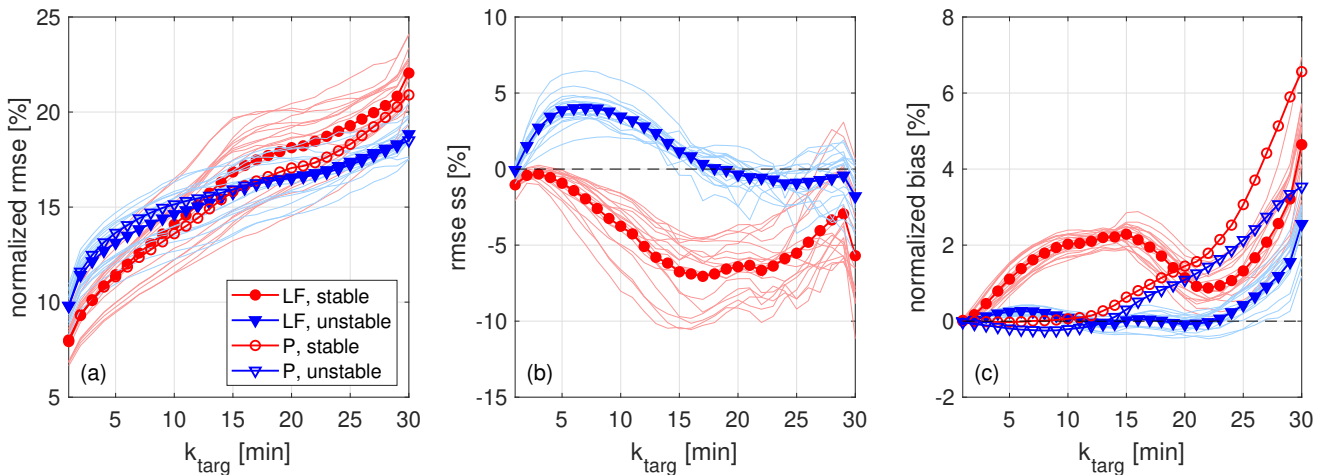
Figure 7. Same as in Figure 6. Here wind speeds that cannot be predicted using the lidar set-up are neglected according to Equation 12.

from previous studies and can be attributed to smaller wind speed fluctuations during stable stratification that can be captured more accurately by both methods. However, during stable cases the rmse increases at a steeper rate and consequently for larger lead times both persistence and the LF are more accurate during unstable situations. We attribute this to changes of larger scale in the wind field which are gaining importance for longer lead times.

275 We compare the skill of the LF with persistence in Figure 6b. Averaged over all turbines of NG, the LF outperforms persistence in terms of rmse for forecast horizons from 2 to 16 minutes for unstable cases, with a maximum rmse ss of 4 % at 7 minutes forecast horizon. For the individual turbines, the maximum rmse ss varies between 2.1 % and 6.5 % and the maximal advantageous forecast horizon lies between 14 minutes and 20 minutes. For stable cases, the LF is not able to outperform persistence. Comparing the rmse and rmse ss of individual turbines, it can be noted that free-stream turbines for main wind 280 direction show lower rmse, i. e. can be predicted more accurately. During unstable cases also the rmse ss is higher for those turbines. In stable cases this is not as distinct as also persistence is more accurate for free stream turbines.

While the bias of persistence lies close to zero for both stable and unstable cases, the LF tends to overestimate wind speed for smaller forecast horizons and underestimates it for larger horizons as visible in Figure 6c. For stable cases, the bias increases to 2% at a 10-minute forecast horizon, decreases until it starts to underestimate wind speed for horizons from 19 minutes 285 onward and reaches a negative value of -3.5% at a 29-minute horizon. The effect is less significant for unstable atmospheric conditions, where the strongest overestimation of 0.3% is observed at a horizon of 6 minutes. The underestimation occurs for horizons from 11 minutes onward and reaches an extreme value of -3% at 29 minutes. For large forecast horizons, wind vectors need to travel a longer duration and consequently longer distances before contributing to a forecast. The limited lidar range cannot capture those distances for high wind speeds, resulting in an underestimation. In agreement with this, a decreasing 290 average wind speed of forecasts with increasing lead times was observed.

To exclude the effect of the limited lidar range on the scores, we neglected wind speeds that cannot be forecasted with the used lidar set-up, i.e. the maximal lidar range. To do so, we assume a lidar range of $r = 12\,\mathrm{km}$ and a wind vector travelling duration of k_{targ}. The maximal wind speed that can be forecasted then yields

$$u_{\mathrm{obs,max}}(t + k_{\mathrm{targ}}) = \frac{r}{k_{\mathrm{targ}}}. \tag{12}$$

295 In Figure 7 the results of this analysis are shown. To analyse the bias in more detail we moreover show the normalized bias binned according to forecast horizon and observed wind speed $u_{\mathrm{obs}}(t + k_{\mathrm{targ}})$ and distinguishing between stable and unstable atmospheric conditions in Figure 8a and b. The cases above the black line are neglected for the analysis shown in Figure 7. In Figure 8c the wind speed distribution for exemplary forecast horizons of 5 minutes, 12 minutes and 20 minutes is depicted for stable and unstable cases. It becomes clear that short forecast horizons are not impacted by the filtering following Equation 12

300 and consequently scores depicted in Figure 7 do not change. For larger forecast horizons, wind speed distributions are already shifted towards lower wind speeds prior to filtering (see Figure 8c). This effect is then further enhanced by applying the filter.

Consequently, the bias of the LF, persistence, stable as well as unstable cases is significantly impacted for lead times of 10 to 15 minutes onwards (see Figure 7). Persistence's bias strongly increases with forecast horizon for stable and unstable cases. Following the definition of persistence, an overestimation is caused by a decrease of wind speed from t to $t + k_{\mathrm{targ}}$. For small

305 observed wind speeds $u_{\mathrm{obs}}(t + k_{\mathrm{targ}})$ it is more likely that strong decreases rather than increases of wind speed have occurred, resulting in the significant overestimation observed for large lead times. Similarly, also the LF's bias increases for large lead times instead of decreasing when considering only wind speeds that the lidar would in principle be able to cover. The changes in bias also affect the rmse and consequently rmse ss. Skill scores are strongly improved for larger horizons when applying the filter. In unstable cases, the rmse ss is close to zero with -0.94 % at a 24 minute horizon. In stable cases skill scores are still

310 significantly negative with -6.7 % at 22 minute horizon, however, individual turbines even reach positive values of up to 3.1 % at 29 minute horizon.

In Figure 8a and b the bias' dependence on wind speed and forecast horizon is visible in more detail. Hereby, two overlapping effects can be identified. Firstly, a with forecast horizon and wind speed increasing negative bias related to the limited lidar range as explained previously. And second, a positive bias for low wind speeds that is slightly increasing with forecast horizon.

315 For low wind speeds, a wind speed reduction from t to $t + k_{\mathrm{targ}}$ is more frequent than an increase. As also the LF utilizes $u_{\mathrm{obs}}(t)$, the effect is similar but less distinct than the one observed for persistence. The higher share of low wind speeds for larger forecast horizons results in the increasing bias with forecast horizon observed in Figure 6 and Figure 7.

3.3 Impact of forecast characteristics on wind speed forecast accuracy

In addition to the dependency on forecast horizon and atmospheric stability, we analysed the impact of forecast characteristics,

320 more specifically the average wind vector age, on forecast skill. Such an analysis can support the interpretation of the LF and can be useful for end-users. Wind speed forecasts were binned according to $\overline{\Delta T}$ in 60 s intervals. Moreover, we distinguished between different forecast horizons and atmospheric stability. Different scores were determined and visualized in Figure 9a-f. For the analysis, forecasts of individual wind turbines were considered.
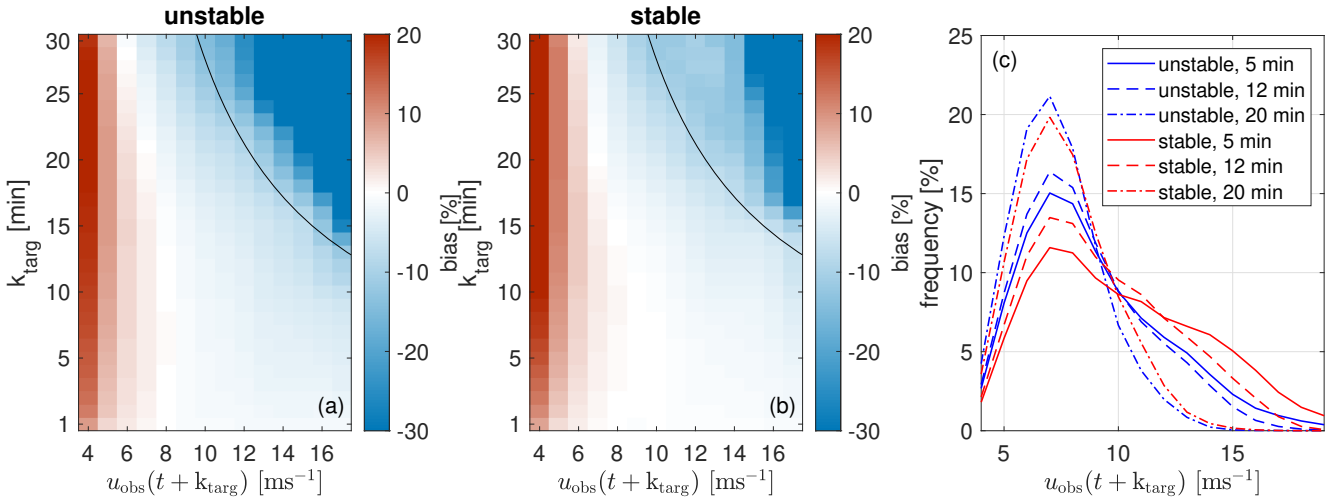
Figure 8. Normalized bias of the wind speed LF binned according to forecast horizon and observed wind speed $u_{\mathrm{obs}}(t + k_{\mathrm{targ}})$ for (a) unstable and (b) stable atmospheric conditions. Black lines indicate the filter criterion according to Equation 12. In (c) the wind speed distribution for exemplary forecast horizons and stable and unstable cases is visualized.

In Figure 9a and d we show the normalized rmse for unstable and stable atmospheric conditions. In agreement with results presented in Section 3.2, forecast errors increase with increasing forecast horizon. Also with growing $\overline{\Delta T}$, the normalized rmse grows for both stable and unstable cases. Both larger forecast horizons and wind vector ages correspond to a longer wind vector travelling duration which is associated with higher uncertainty. One should note, however, that the forecast horizon has an additional effect on the forecast skill as it impacts also other variables contributing to the forecast such as meteorological variables and SCADA data.

An exception to the increasing average rmse with increasing $\overline{\Delta T}$ are very low vector ages and up to medium range forecast horizons for which the opposite holds. While most combinations of k_{targ} and $\overline{\Delta T}$ are dominated by a negative bias, for these cases an overestimation of wind speed was observed as shown in Figure 9c and f. In particular for stable conditions, a distinct positive bias is visible for average vector ages of up to 10 minutes and forecast horizons up to 20 minutes.

Small $\overline{\Delta T}$ result from wind speed forecasts for which most wind vectors are selected from very recent or, in the most extreme case, only the most recent lidar scan. In Figure 10 the dependence of the number of individual lidar scans contributing to a forecast and the average scan ID of the wind vectors contributing, with lower IDs corresponding to more recent scans, on the average wind vector age is shown. Results are averages for all wind turbines and forecast horizons with error bars indicating their standard deviation. The figure clearly shows that small vector ages correspond to a low number of scans considered. As a consequence of the wind vector age's definition, for those cases mainly the most recent scans contribute. We will discuss the impact of lidar trajectories and wind vector propagation on the forecast's bias in more detail in Section 4.2.

The rmse ss in Figure 9b and e shows that the LF is able to outperform persistence for forecast horizons up to 19 minutes for unstable cases when the average wind vector age lies below 2 minutes. This analysis also revealed that even during stable
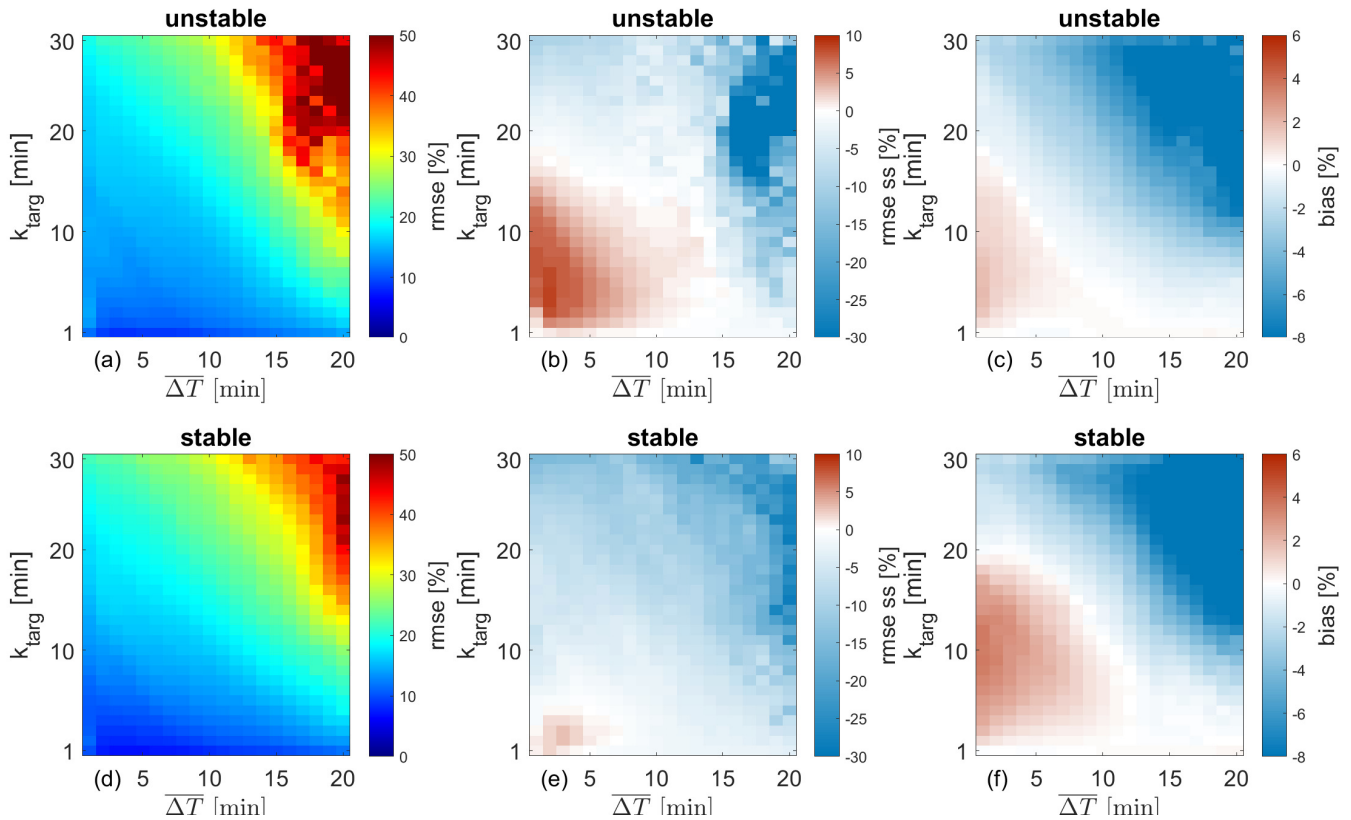
Figure 9. Wind speed forecast skill dependent on forecast horizon and average wind vector age and distinguishing between unstable atmospheric conditions (upper row) and stable atmospheric conditions (lower row). In (a) and (d) the normalized rmse in %, in (b) and (e) the rmse skill score in % with persistence as reference and in (c) and (f) the normalized bias in % are shown.

stratification persistence can be outperformed for lead times up to 5 minutes and average wind vector ages between 2 and 5 minutes. The highest skill score of 8.6 % was achieved for a forecast horizon of 5 minutes and an average wind vector age of 2 minutes during unstable stratification.

3.4 Evaluation of individual turbine and wind farm power forecasts

After the detailed analysis of the wind speed forecast's skill in the previous sections, we will now focus on the evaluation of power forecasts. As for many applications, not individual turbine but wind farm power is of importance, this analysis will include wind farm power forecasts. The Figures 11a, b and c visualize the same scores as Figure 6. Here we compare the average of the normalized rmse, rmse ss and bias of individual turbines of Nordergründe.

The results align well with the wind speed forecasts presented in Section 3.2. The behaviour of normalized rmse and bias is for both LF and persistence and during stable as well as unstable cases very similar to that observed for the wind speed forecast.
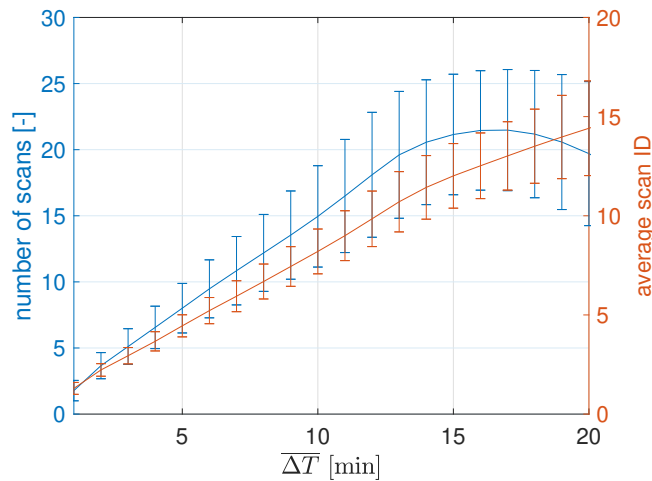
Figure 10. The number of individual lidar scans contributing to a LF and the average scan ID of the contributing wind vectors dependent on the average wind vector age is shown. Markers indicate the average values for all wind turbines and forecast horizons and error bars the corresponding standard deviation.

In general, scores are higher in this case due to the cubic nature of the power curve. The LF is able to outperform persistence during unstable cases for forecast horizons up to 17 minutes. The maximal rmse ss yields 6.4 % at 9-minute forecast horizon

355   with maximal values for the individual turbines ranging between 4.2 % and 9.8 %. During stable cases persistence outperforms the LF.

In Figure 12 we evaluate the wind farm power forecast. Following the analysis in Section 3.3, in addition to the atmospheric condition we also consider the average propagation duration here. The average propagation duration is defined as the average wind vector age plus the forecast horizon. We show five different cases, namely i) all valid forecasts as evaluated in Figure 11,

360   ii) forecasts during unstable atmospheric conditions, iii) forecasts during unstable conditions and with an average propagation duration below 25 minutes, iv) below 20 minutes and v) below 16 minutes. For all cases, the rmse ss with persistence as reference is depicted. The 95 % confidence intervals were generated using bootstrapping with replacement and are depicted as error bars. Additionally, the number of valid forecasts at a 5 minute and a 15 minute forecast horizon, $N_{5\,\mathrm{min}}$ and $N_{15\,\mathrm{min}}$, normalized with $N_{5\,\mathrm{min}}$ of case i), the maximal rmse ss and its corresponding lead time, $\mathrm{rmse\,ss_{max}}$ and $k_{\mathrm{targ}}(\mathrm{rmse\,ss_{max}})$, the

365   rmse ss at 15 minute forecast horizon $\mathrm{rmse\,ss_{15\,min}}$ and the maximal lead time with a positive rmse ss $k_{\mathrm{targ,max}}(\mathrm{rmse\,ss} > 0)$ are summarized in Table 3. The skill at a 15-minute horizon holds significant value when aiming at the application of the forecast for trading and system operation.

Compared to the individual turbines' power forecasts the rmse skill score is larger for wind farm power forecasts with a maximum value of 12.1 % at a forecast horizon of 6 minutes. During unstable conditions, the rmse ss even reaches values

370   as high as 19.0 % at a 7-minute forecast horizon. Through the averaging, forecast errors of individual turbines average out, resulting in an overall higher forecast skill. Though the same applies for persistence, the effect seems to be stronger for the
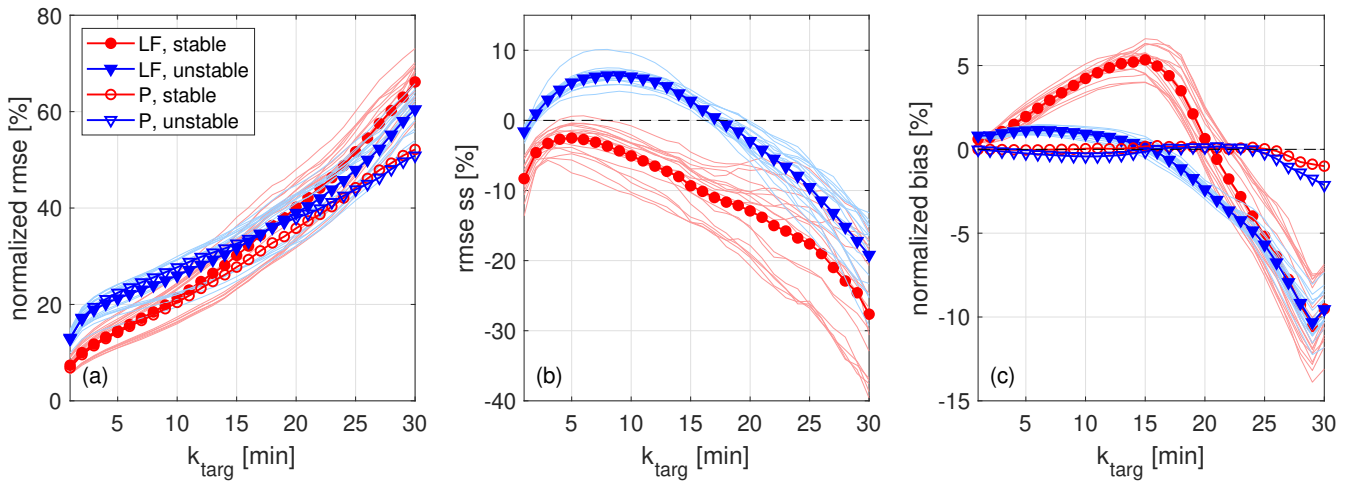
Figure 11. (a) Normalized rmse and (c) bias of the power LF (filled markers) and persistence (empty markers) for forecast horizons up to 30 minutes and distinguishing between stable cases as red circles and unstable cases as blue triangles. Thick lines with markers indicate the turbines' average scores while thin lines without markers show scores of individual turbines. In (b) the rmse skill score in % is shown. Black dashed lines indicate a 0 % skill score and normalized bias, respectively.
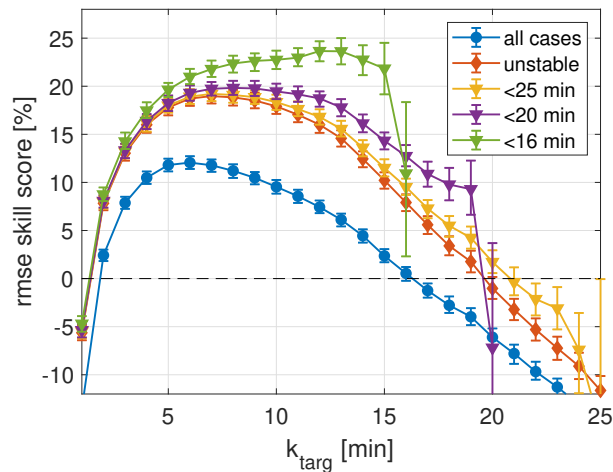
Figure 12. The rmse ss with persistence as a reference for all available wind farm power forecasts, forecasts during unstable situations and forecasts during unstable cases with an average propagation duration below 25 min, 20 min and 15 min respectively. Error bars indicate 95 % confidence intervals. The black dotted line indicates a skill score of 0 %.

Table 3. The number of valid forecasts at 5 minute forecast horizon $N_{5\,\mathrm{min}}$ and at 15 minute forecast horizon $N_{15\,\mathrm{min}}$ normalized with $N_{5\,\mathrm{min}}$ of case i), the maximal rmse ss $\mathrm{rmse\,ss_{max}}$ and its corresponding lead time $k_{\mathrm{targ}}(\mathrm{rmse\,ss_{max}})$, the rmse ss at 15 minute forecast horizon $\mathrm{rmse\,ss_{15\,min}}$ and the maximal lead time with a positive rmse ss $k_{\mathrm{targ,max}}(\mathrm{rmse\,ss} > 0)$ for all data and four different filters applied.

	$N_{5\,\mathrm{min}}$ [%]	$N_{15\,\mathrm{min}}$ [%]	$\mathrm{rmse\,ss_{max}}$ [%]	$k_{\mathrm{targ}}(\mathrm{rmse\,ss_{max}})$ [min]	$\mathrm{rmse\,ss_{15\,min}}$ [%]	$k_{\mathrm{targ,max}}(\mathrm{rmse\,ss} > 0)$ [min]
i) all	100	73.51	13.92	6	2.36	16
ii) unstable	61.51	47.91	19.94	7	10.21	19
iii) unstable & $\Delta T_{\mathrm{prop}} < 25\,\mathrm{min}$	60.75	40.15	20.17	7	11.50	20
iv) unstable & $\Delta T_{\mathrm{prop}} < 20\,\mathrm{min}$	54.04	24.07	20.73	8	14.33	19
v) unstable & $\Delta T_{\mathrm{prop}} < 16\,\mathrm{min}$	41.02	2.95	27.66	12	21.84	16

LF, resulting in a larger rmse ss. As shown in Section 3.3, large wind vector ages and large forecast horizons are associated with higher uncertainty. By excluding those cases, the rmse ss can be improved, in particular for larger lead times for which the confidence intervals confirm the significance of the results. Also the maximal forecast horizon with a positive skill score, thus outperforming persistence, increases for unstable conditions when considering an average propagation duration below 25 minutes (case iii)) in comparison to all available forecasts (case i)) and unstable situations (case ii)). For unstable cases with an average propagation duration below 20 minutes (case iv)) and 15 minutes (case v)) it decreases, along with a significant drop in the number of available forecasts, particularly for larger lead times.

4 Discussion

In this section we discuss the accuracy and feasibility of the wind speed extrapolation approaches introduced in this work (Section 4.1), the impact of lidar trajectories and wind vector propagation on forecast skill (Section 4.2) and the finally the possible value of the introduced minute-scale power forecast for end-users (Section 4.3).

4.1 Wind speed extrapolation approaches and lidar tilt correction

For the determination of wind power forecasts, the wind speed extrapolation from measurement to hub height is a crucial part of the forecasting methodology. As discussed in various previous studies a lidar tilt can cause significant differences in measurement height when considering the large measurement distances of the utilized long-range devices (Theuer et al., 2020b, 2021; Schneemann et al., 2021). Previous studies have shown that applying a stability-corrected logarithmic profile for wind speed extrapolation can cause large errors in particular during stable atmospheric conditions and for profiles deviating from the assumed logarithmic shape such as low level jets or reversed profiles (Theuer et al., 2020a, 2021). While the forecasting method applied here, referred to as tendency method in previous work, was proven to be less sensitive to these impacts, a dependence of rmse on the extrapolation distance was nonetheless proven by Theuer et al. (2021). This indicates that further development of extrapolation methodologies is worthwhile.

In good agreement with these expectations, our analysis has shown that all applied extrapolation approaches, i. e. measured or logarithmic wind profiles without tilt correction or with a correction based on a tilt model or inclination sensors, were
395 able to improve forecast skill. However, after reaching a maximum value at medium-range forecast horizons, skill scores start to decrease and even reach negative values for larger lead times. In those cases, the forecast uncertainty is not dominated by the height correction and wind speed extrapolation but by other impacts, such as wind vector propagation, reducing their impact on the overall forecast skill. At the same time, the height differences that need to be corrected for are increasing with forecast horizon as they are mainly related to wind speed changes (Theuer et al., 2021), which are expected to increase with
400 time increments. Thus, improving profiles has a stronger impact for larger forecast horizons, explaining the increasing rmse ss observed up to a 15 minute forecast horizon. On the other hand, we attribute the negative skill scores when using logarithmic profiles for larger forecast horizons to a similar effect, i.e. the application of inadequate profiles.

Utilizing wind profile measurements has proven advantageous over the use of stability-corrected logarithmic wind profiles (Møller et al., 2020) and aligns well with the results from Theuer et al. (2021) mentioned above. For further work, an analysis
405 of the impact of these measured profiles on forecast skill during e.g. low-level jets might be interesting. Furthermore, the extraction of wind profiles might be useful to determine a rotor equivalent wind speed from single-value wind speed forecasts at hub height and based on that more accurate power forecasts. This might become in particularly relevant for future offshore wind turbines with very large rotor diameters. In this work we used a set of very fast PPI scans with varying elevation angles to determine the wind profile. This allowed to determine highly resolved wind profiles in the inflow wind field and did not require
410 additional measurement equipment. However, it also reduced the amount of horizontal scans that could be propagated and put additional strain on the lidar devices, especially considering the fast scanning speed. For future campaigns, we thus suggest to rely on an additional lidar device that measures wind profiles in VAD mode at a location close to the location of the PPI lidar. Alternatively, scanning trajectories that combine wind vector propagation and profile estimation could be explored. That means, a set of scans with few different elevations and lower scanning speed that cover relatively large ranges and azimuth
415 sectors. Such an approach could also be beneficial to determine height-resolved wind speed forecasts.

The estimation of the measurement height using the tilt model was similarly accurate to using the inclination sensor data. One should note, however, that the calibration of the tilt model is based on the availability of inclination data, determined either by inclination sensors or so-called sea surface levelling as described in Section 2 and Rott et al. (2022). Moreover, high frequency SCADA data needs to be available. That means, the tilt model cannot be applied directly after deploying the lidar
420 but a certain amount of data needs to be collected to be able to calibrate it. To analyse how much data is required was out of the scope of this work, however, in general it can be assumed that longer periods covering a larger variety of conditions lead to better results.

Overall, our results support the notion that the tilt model can be used as a feasible alternative to inclination sensor data when sensors fail during the campaign. The same holds for the use of stability-corrected logarithmic profiles as an alternative
425 to measured profiles, however, to a lesser extent. The identification of viable fall-back options is important when aiming at generating a fail-safe forecast for real-time applications and can thus be considered an important step for pushing the lidar-based forecasts towards an operational use.

4.2 Relation of forecast skill and horizon

In our analysis we have shown that a longer wind vector propagation duration, i.e. a larger forecast horizon or wind vector age, corresponds to larger forecast uncertainties. Besides higher uncertainty associated to the vector propagation, this is also linked to the forecast's negative bias. This agrees well with results presented by Würth (2022) who observed an increasing underestimation of wind speed and power for forecast horizons from 10 to 30 minutes for both an onshore and an offshore site. High wind speeds travel longer distances in the same duration and can thus not be captured adequately by the maximum measured distance. Excluding those wind speeds from the analysis strongly improved the results and removed the underestimation. This effect is strongly related to the maximal range of the lidar, but the selected trajectories, wind direction and wind farm layout also play an important role. While we tried to optimize the scanned area (Section 2.1), the selected trajectories are still a trade-off between scanning duration, the ability to cover all turbines of the wind farm adequately and a sufficient azimuth angle spanned for robust VAD-fitting.

We have already discussed the systematic underestimation of wind speed related to the measurement range. This effect might be enhanced by the utilization of several subsequent lidar scans: As low wind speeds travel slower, they will be captured in more of the considered lidar scans and can thus repeatedly contribute to the forecast. The time window within which wind vectors are selected to contribute has the opposite effect. A broader range of distances is eligible for higher wind speeds, which could result in an overestimation. These considerations agree well with the positive bias observed for low average wind vector ages which were shown to relate to a low number of more recent scans considered. If only few scans contribute to a forecast, the effect of the time window will dominate over that of the slow vectors' appearance throughout more scans. While the selection of only few scans can in principle be related to the nature of the wind field, the effect might be enhanced during data gaps when the lidar was not measuring or caused by low-quality scans. As described in Section 2.2, these scans were not considered for forecasting. The positive bias might be enhanced by slow wind vectors that would originate close to the turbines but are not captured by the most recent lidar scan as a consequence of the scanning trajectory and wind farm layout. Theuer et al. (2020a) discuss a similar effect in their work that they attribute to an area not covered well by the lidar scans.

When analysing the bias, one needs to consider that the impact of systematic errors in e.g. wind vector propagation is reduced by using the tendency approach (Theuer et al., 2021). That means, the bias observed during forecast evaluation is caused by differences in the bias of forecasts with k_{tend} and k_{targ} as described in Section 2.2. With $k_{\mathrm{tend}} < k_{\mathrm{targ}}$ and Equation 5 this results in an increasing underestimation of wind speed with larger lead times. For $k_{\mathrm{targ}} < 14\,\mathrm{min}$ the difference between k_{tend} and k_{targ} increases before reaching a constant value of 13 min (see Equation 2). Consequently, also the difference in bias increases, which we assume contributes to the increasing positive bias observed during the analysis. Moreover, we were able to identify the changing wind speed distribution towards a higher share of low wind speeds with increasing forecast horizon as a main reason for the increasingly positive bias. Apart from the wind speed tendency, also the currently observed wind speed contributes strongly to the LF. As also observed for persistence, low wind speeds at $t + k_{\mathrm{targ}}$ are more often related to negative wind speed increments from t to $t + k_{\mathrm{targ}}$, resulting in an overestimation of wind speed. While for persistence the opposite

holds for larger wind speeds, resulting in an overall bias close to 0, for the LF the overestimation dominates, in particular for stable cases.

To reduce the systematic error caused by the utilization of subsequent lidar scans, a temporal weighting could be introduced. Rott et al. (2020) use spatio-temporal weighting in their work to account for the spatial distance of wind vectors to the target

465 turbine and the increasing uncertainty of the assumptions related to Lagrangian advection. In Theuer et al. (2022a) a temporal weighting was applied to a SCADA-based forecast that was then combined with a lidar-based approach to an observer-based forecast. Considering the generally increasing uncertainty related to wind vector advection with propagation duration, a temporal weighting of wind vectors is considered promising and should be explored further. The additional use of SCADA data can also reduce data gaps caused by the lidar trajectory, in particularly close to the target wind turbine, which might reduce

470 the positive bias. It would also be interesting to analyse how the scanning duration and thus repetition rate of individual lidar scans impacts the results. Moreover, further work could explore the possibility of scaling the time window during which wind vectors are selected with wind speed.

4.3 Operational use of minute-scale forecasts

As mentioned previously, forecast horizons of at least 15 minutes with an additional processing time of few minutes are

475 required for application in electricity trading and system operation. For that purpose, the evaluation of wind farm power forecasts is particularly important, while the evaluation of wind speed forecast at individual turbines is crucial to understand the forecast and its limitations. While our results show that the lidar-based forecast is not able to outperform persistence for all situations and forecast horizons it clearly has the potential to do so under specific conditions. A main influencing factor is atmospheric stability. In previous studies from Theuer et al. (2021) and Theuer et al. (2022a) the authors showed that the LF can

480 outperform persistence during unstable atmospheric conditions in good agreement with this work. Also Pichault et al. (2021b) found that their lidar-based wind farm power forecast outperformed persistence during unstable and particularly neutral cases, however, not during stable ones. The authors also identified a dependence on wind direction which they attribute to differences in measurement height, terrain and variability in the wind associated to the different wind directions. Valldecabres et al. (2018b) related the higher forecast skill of their lidar-based wind speed forecast during neutral situations as compared to stable ones to

485 the logarithmic wind profile without stability correction that they applied for height correction. Moreover, Valldecabres et al. (2020), Würth et al. (2018) and Pichault et al. (2021b) found their remote sensing-based approaches to outperform persistence in particular during ramp events. A thorough analysis of an individual turbine's probabilistic wind speed and power forecast at an offshore location from 0 minutes to 10 minutes forecast horizon by Würth (2022) revealed a dependence of forecast skill on wind speed. The author related this to neglecting the lidar tilt, that results in a difference between the measurement

490 height and hub height and causes larger absolute errors for higher wind speeds. Regarding the forecast characteristics, Würth (2022) found an impact of the number of wind vectors contributing to a forecast and the distance those vectors originate from. While considering more wind vectors is related to smaller errors, larger distances correspond to larger ones. For operational use, the distinction between favourable and unfavourable LFs as compared to a reference method is important. What has proven especially useful in our analysis is the consideration of atmospheric stability and the propagation duration for forecast

495 evaluation. In the results shown in Section 3.4, the LF was able to outperform persistence for horizons up to 20 minutes and for a 15-minute horizon with skill scores of 11.9 % when only unstable cases and a propagation duration below 25 minutes were considered. The LF can thus be considered a valuable extension to statistical methods. It might be rewarding to explore more sophisticated approaches to select the beneficial LFs than the one presented in this study, e.g. the consideration of more parameters characterizing the forecast or the implementation of machine learning approaches. Also the development of a hybrid

500 approach could further improve the results. For instance, Theuer et al. (2022b) showed that an observer-based forecast can support persistence also for situations during which it was not able to outperform it. Our results also confirmed the dependence of forecast skill on turbine position that was already discussed by Theuer et al. (2022b). Free-stream turbines can be predicted more accurately than inner-farm ones. Further work could analyse to what extent power forecasts of free-stream turbines could be extrapolated to wind farm power forecasts, neglecting less accurate individual turbine forecasts.

505 We expect the circumstances under which the LF is able to outperform persistence to be related to the specific wind farm layout, conditions at the site (e.g. coastal effects) and scanned trajectories. Thus, a detailed analysis and a sufficiently large data set to train hybrid models or selection algorithms would be required. It should be noted that a strict selection of favourable LFs might improve its skill score as compared to a reference method, however, it also reduces the number of selected cases, i.e. the number of situations during which the lidar forecast is applied. A cost-benefit analysis for the different areas of application is

510 required to assess the value of the LF under these circumstances. However, several studies have shown the value of lidar-based approaches in particular during ramp events (Valldecabres et al., 2020; Pichault et al., 2021b; Würth, 2022). As ill-predicted ramp events can become very costly both in electricity trading and grid operation, we expect the lidar-based forecast to be cost-effective despite the ramp events' relatively low frequency of occurrence. Due to brevity, we did not analyse ramp events in more detail in this work but aim to do so in a future study.

515 ## 5 Conclusions

In this work we extended the forecast horizon of a minute-scale wind speed and power forecast from 5 to 30 minutes utilizing long-range lidar data and adapting a lidar-based forecasting methodology. An important contributor to increasing the forecast's horizon was carefully planning the offshore measurement campaign, optimizing lidar trajectories and installing complementary sensors, i.e. meteorological sensors and inclination sensors. In this context, applying a model for the static and operation-

520 induced mean tilt was proven a valid alternative to using inclination sensors to determine the lidar's tilt. Further, stability corrected logarithmic profiles can be used to correct measurement heights, however, using wind profiles measured by multi-elevation lidar scans is favourable. Moreover, and in good agreement with previous studies, we have found the forecast skill, more specifically the forecast's bias, to be impacted by the lidar trajectories, i.e. the area covered by the scans. In this context, particularly the measurement range plays an important role.

525 For a medium-sized offshore reference wind farm we showed that the lidar-based approach is able to outperform the statistical benchmark persistence for unstable atmospheric conditions up to forecast horizons of 16 minutes for individual turbine wind speed and power forecasts. During stable conditions the lidar-based approach was not able to outperform persistence.

Our results further revealed the dependence of the forecast skill on the average wind vector age and propagation duration. Considering these forecast characteristics has proven useful for the assessment of wind farm power forecasts and can be beneficial to end-users. Overall, we conclude that lidar-based approaches can contribute substantially to generating wind farm power forecasts. In future work, their application in electricity trading and system operation needs to be further developed and tested, including, among others, the prediction of ramp events and their use complementary to other methods, i.e. hybrid approaches that combine e.g. statistical methods and the lidar-based approach.

Data availability. Lidar and meteorological data are not published and could be made available on request. The OSTIA data set can be obtained from http://marine.copernicus.eu (Copernicus marine service, 2022). NG SCADA data is confidential and therefore not available to the public.

Author contributions. FT conducted the main research and wrote the manuscript. JS executed the measurement campaign, supported the lidar data analysis, contributed to the scientific discussion and provided extensive feedback in form of manuscript reviews. JKS processed the meteorological data and SCADA data, contributed to the scientific discussion and extensively reviewed the manuscript. MK supervised the work, contributed to the scientific discussion and reviewed the manuscript.

Competing interests. The authors declare no conflict of interests.

Acknowledgements. The lidar measurements and the work were performed within the research projects WindRamp (Ref. Nr. 03EE3027A) funded by the German Federal Ministry for Economic Affairs and Climate Action on the basis of a decision by the German Bundestag. We acknowledge the wind farm operator Skyborn Renewables GmbH for providing SCADA data and thank them for supporting the measurement campaign in NG and our work. We acknowledge the UK Met Office for making the OSTIA data set available. We thank Timo Brunßen, Daniel Ribnitzky and Lilén A. Vöhringer for their support conducting the measurement campaign, Johannes Paulsen for providing the wind profiles and Andreas Rott for reviewing the manuscript.

References

Copernicus marine service: Copernicus Marine environment monitoring service, available at: http://marine.copernicus.eu/, last access: 05 January 2024, 2022.

European Commission: European Green Deal: EU agrees stronger legislation to accelerate the rollout of renewable energy, available at: https://ec.europa.eu/commission/presscorner/detail/en/IP_23_2061, last access: 05 January 2024, 2023.

Good, S., Fiedler, E., Mao, C., Martin, M. J., Maycock, A., Reid, R., Roberts-Jones, J., Searle, T., Waters, J., While, J., and Worsfold, M.: The Current Configuration of the OSTIA System for Operational Production of Foundation Sea Surface Temperature and Ice Concentration Analyses, Remote Sensing, 12, 720, https://doi.org/10.3390/rs12040720, 2020.

IEC: IEC 61400-12-1:2017 Wind energy generation systems - Part 12-1: Power performance measurements of electricity producing wind turbines, ed. 2, 2017.

Møller, M., Domagalski, P., and Sætran, L. R.: Comparing Abnormalities in Onshore and Offshore Vertical Wind Profiles, Wind Energy Science, 5, 391–411, https://doi.org/10.5194/wes-5-391-2020, 2020.

Paulsen, J., Schneemann, J., Steinfeld, G., Theuer, F., and Kühn, M.: Low-level jet's influence on offshore wind turbine's power production, in preparation for Wind Energy Science, 2024.

Pichault, M., Vincent, C., Skidmore, G., and Monty, J.: Characterisation of intra-hourly wind power ramps at the wind farm scale and associated processes, Wind Energy Science, 6, 131–147, https://doi.org/10.5194/wes-6-131-2021, 2021a.

Pichault, M., Vincent, C., Skidmore, G., and Monty, J.: Short-Term Wind Power Forecasting at the Wind Farm Scale Using Long-Range Doppler LiDAR, Energies, 14, 2663, https://doi.org/10.3390/en14092663, 2021b.

Rott, A., Petrović, V., and Kühn, M.: Wind farm flow reconstruction and prediction from high frequency SCADA Data, Journal of Physics: Conference Series, 1618, 062 067, https://doi.org/10.1088/1742-6596/1618/6/062067, 2020.

Rott, A., Schneemann, J., Theuer, F., Trujillo Quintero, J. J., and Kühn, M.: Alignment of scanning lidars in offshore wind farms, Wind Energy Science, 7, 283–297, https://doi.org/10.5194/wes-7-283-2022, 2022.

Schneemann, J., Theuer, F., Rott, A., Dörenkämper, M., and Kühn, M.: Offshore wind farm global blockage measured with scanning lidar, Wind Energy Science, 6, 521–538, https://doi.org/10.5194/wes-6-521-2021, 2021.

Sweeney, C., Bessa, R. J., Browell, J., and Pinson, P.: The future of forecasting for renewable energy, WIREs Energy and Environment, 9, e365, https://doi.org/10.1002/wene.365, 2020.

Theuer, F., van Dooren, M. F., von Bremen, L., and Kühn, M.: On the accuracy of a logarithmic extrapolation of the wind speed measured by horizontal lidar scans, Journal of Physics: Conference Series, 1618, 032 043, https://doi.org/10.1088/1742-6596/1618/3/032043, 2020a.

Theuer, F., van Dooren, M. F., von Bremen, L., and Kühn, M.: Minute-scale power forecast of offshore wind turbines using single-Doppler long-range lidar measurements, Wind Energy Science, 5, 1449–1468, https://doi.org/10.5194/wes-5-1449-2020, 2020b.

Theuer, F., van Dooren, M. F., von Bremen, L., and Kühn, M.: Lidar-based minute-scale offshore wind speed forecasts analysed under different atmospheric conditions, Meteorologische Zeitschrift, 31, 13–29, https://doi.org/10.1127/metz/2021/1080, 2021.

Theuer, F., Rott, A., Schneemann, J., von Bremen, L., and Kühn, M.: Observer-based power forecast of individual and aggregated offshore wind turbines, Wind Energy Science, 7, 2099–2116, https://doi.org/10.5194/wes-7-2099-2022, 2022a.

Theuer, F., Schneemann, J., van Dooren, M. F., von Bremen, L., and Kühn, M.: Hybrid use of an observer-based minute-scale power forecast and persistence, Journal of Physics: Conference Series, accepted, 2265, 022 047, https://doi.org/10.1088/1742-6596/2265/2/022047, 2022b.

585 Thorarinsdottir, T. L. and Gneiting, T.: Probabilistic forecasts of wind speed: ensemble model output statistics by using heteroscedastic censored regression, Journal of the Royal Statistical Society: Series A (Statistics in Society), 173, 371–388, https://doi.org/10.1111/j.1467-985X.2009.00616.x, 2010.

Ulazia, A., Nafarrate, A., Ibarra-Berastegi, G., Sáenz, J., and Carreno-Madinabeitia, S.: The Consequences of Air Density Variations over Northeastern Scotland for Offshore Wind Energy Potential, Energies, 12, 2635, https://doi.org/10.3390/en12132635, 2019.

590 Valldecabres, L., Nygaard, N., Vera-Tudela, L., von Bremen, L., and Kühn, M.: On the Use of Dual-Doppler Radar Measurements for Very Short-Term Wind Power Forecasts, Remote Sensing, 10, 1701, https://doi.org/10.3390/rs10111701, 2018a.

Valldecabres, L., Peña, A., Courtney, M., von Bremen, L., and Kühn, M.: Very short-term forecast of near-coastal flow using scanning lidars, Wind Energy Science, 3, 313–327, https://doi.org/10.5194/wes-3-313-2018, 2018b.

Valldecabres, L., von Bremen, L., and Kühn, M.: Minute-Scale Detection and Probabilistic Prediction of Offshore Wind Turbine Power
595 Ramps using Dual-Doppler Radar, Wind Energy, 23, 1–23, https://doi.org/10.1002/we.2553, 2020.

Werner, C.: Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere, chap. 12 - Doppler Wind Lidar, pp. 325–354, Springer New York, New York, NY, https://doi.org/10.1007/0-387-25101-4_12, 2005.

WindEurope: Wind energy in Europe - 2022 Statistics and the outlook for 2023-2027, Tech. rep., 2023.

Würth, I.: Minute-scale forecasting of wind power using long-range lidar data, Ph.D. thesis, University of Stuttgart, 2022.

600 Würth, I., Ellinghaus, S., Wigger, M., J Niemeier, M., Clifton, A., and W Cheng, P.: Forecasting wind ramps: can long-range lidar increase accuracy?, Journal of Physics: Conference Series, 1102, 012 013, https://doi.org/10.1088/1742-6596/1102/1/012013, 2018.

Würth, I., Valldecabres, L., Simon, E., Möhrlen, C., Uzunoğlu, B., Gilbert, C., Giebel, G., Schlipf, D., and Kaifel, A.: Minute-Scale Forecasting of Wind Power – Results from the Collaborative Workshop of IEA Wind Task 32 and 36, Energies, 12, 712, https://doi.org/10.3390/en12040712, 2019.