Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Hybrid use of an observer-based minute-scale power forecast and persistence

Hybrid use of an observer-based minute-scale power forecast and persistence

F Theuer[1], J Schneemann[1], M F van Dooren[1], L von Bremen[2] and M Kühn[1]

[1]ForWind - Center for Wind Energy Research, Institute of Physics, University of Oldenburg, Küpkersweg 70, 26129 Oldenburg, Germany

[2]DLR Institute of Networked Energy Systems, Carl-von-Ossietzky-Str. 15, 26129 Oldenburg, Germany

E-mail: `frauke.theuer@uni-oldenburg.de`

Abstract. Lidar-based minute-scale wind power forecasts are valuable to support grid stability and electricity trading. Current methodologies are able to outperform the benchmark persistence only during transient situations and unstable stratification. So far, methods that extend lidar-based forecasts to observer-based forecasts by embedding turbine operational data are not able to outperform persistence during stable atmospheric conditions either. In this paper we therefore analyse the complementary use of an observer-based power forecast and persistence. To do so, we implemented two hybrid approaches: The first is based on a binary decision algorithm, while the second is weighting the two methods by minimizing a cost function. We evaluated 5-minute-ahead deterministic power forecasts of the hybrid and individual models at an offshore wind farm and found the weighting approach to be most skillful. Further, the data set was extended to represent the atmospheric conditions on site for an entire typical year. The weighting approach outperformed the binary decision algorithm for both the 5-minute sample forecasts and the one year-long data set. We discuss the advantages and disadvantages of the two hybrid models and conclude that the weighting approach is the better choice. Further, it can be concluded that also when evaluating the forecasts over a longer period, in this case one year, the additional use of observer-based forecasts is beneficial compared to solely relying on persistence.

1. Introduction

In the context of the increased share of renewable energies in our power system, minute-scale power forecasts are developed to support grid integration and electricity trading [1]. On these time horizons, typically statistical methods are used. A highly competitive and easy-to-implement statistical benchmark is persistence, which assumes that the current value equals the future one [1]. Probabilistic extensions of persistence consider past forecasting errors to build a forecast distribution [2]. Other time series models are, for instance, AR(I)MA (autoregressive (integrated) moving average) models that utilize a number of past measurements and in some cases forecasting errors [1]. It is common to combine different forecasting models to a hybrid approach in order to increase the available information contained in the forecast, exploit the advantages of both individual methods and thereby improve forecast skill [3]. Hybrid models can combine different types of forecasts, i. e. statistical and physical models, or forecasts with different lead times.

In the last years, remote sensing-based minute-scale wind speed and power forecasts have been the subject of research as an alternative to statistical methods [4][5][6][7]. Valldecabres et al. developed a promising radar-based approach that was able to outperform persistence for free-stream turbines and during ramp events [8, 4]. Studies have shown that lidar-based forecasts (LF) are able to outperform persistence during specific situations, namely during unstable atmospheric conditions and situations with high turbulence intensity or large wind speed increments. However, also lower forecast skills were observed for other conditions [6]. A recent study that has extended lidar-based forecasts to observer-based forecasts (OF) by embedding also wind turbine operational data confirmed the superiority of persistence in particular during stable stratification [9, 10]. This suggests using observer-based methods as a complement to persistence rather than a substitute. It also implies that the benefit of OFs evaluated over longer time periods will strongly depend on the atmospheric conditions and stratification at the wind farm location.

The aim of this work is to gain a more comprehensive understanding of the benefit of observer-based power forecasts as a supplement to persistence. We develop strategies to apply the OF in combination with persistence under different atmospheric conditions. We evaluate the forecast skill for a training data set and additionally for an artificial one year-long data set, that mimicks the atmospheric conditions at the test site.

2. Methodology

In this section we first introduce the observer-based power forecast that utilizes both lidar and SCADA (Supervisory Control and Data Acquisition) data. Distinguishing between atmospheric stability, the power production and wake impact, two hybrid methods combining persistence and the observer-based forecast are then introduced. The first approach is based on a binary decision algorithm and the second method weights and combines the two individual forecasts. Finally, we explain the generation of artifical one year-long OFs, hybrid forecasts and persistence forecasts by resampling from the original forecasts in order to assess the methods' usefulness over a longer time period. Hereby, the artifical data set mimicks the typical atmospheric conditions at the wind farm site.

2.1. Minute-scale observer-based power forecasts at the offshore wind farm Global Tech I

The 5-minute ahead lidar-based forecasts were generated for a period from March 2019 until June 2019 using horizontal plan position indicator (PPI) lidar scans performed at the offshore wind farm Global Tech I (GT I). The wind farm consists of 80 turbines of type Adwen AD 5-116 with a rated power P_r of 5 MW. Lidar scans were measured with an azimuthal resolution of 2°, an opening angle of 150°, range gates from 500 m to 7950 m in 35 m steps and an accumulation time of 2 s. The scan orientation was adjusted manually to one of four sectors according to wind direction. Figure 1 depicts the wind farm layout and scanning trajectories. Line-of-sight (LOS) wind speed measurements were transformed to horizontal wind speed using wind direction information estimated by means of a Velocity Azimuth Display (VAD) fit individually for each range gate [11]. The resulting wind vectors were propagated using Lagrangian advection [8]. Those vectors reaching an area of influence around the target turbine within the time interval $k \pm 30$ s, with lead time k, were selected to contribute to the wind speed forecast. Two subsequent wind speed forecasts at lidar height, one initialized at t and one at $t-k$, were used to determine a wind speed tendency. In a next step, this tendency was applied to high-frequency operational wind turbine data to determine a wind speed forecast at hub height.

Here, we additionally propagated high-frequency wind speed and wind direction information obtained from GT I's wind turbine operational data as suggested by [9]. Wind vectors based on SCADA data were weighted according to their age and bias-corrected to account for wake effects [10]. Lidar and SCADA contributions were resampled to the same number of wind

Journal of Physics: Conference Series **2265** (2022) 022047 doi:10.1088/1742-6596/2265/2/022047

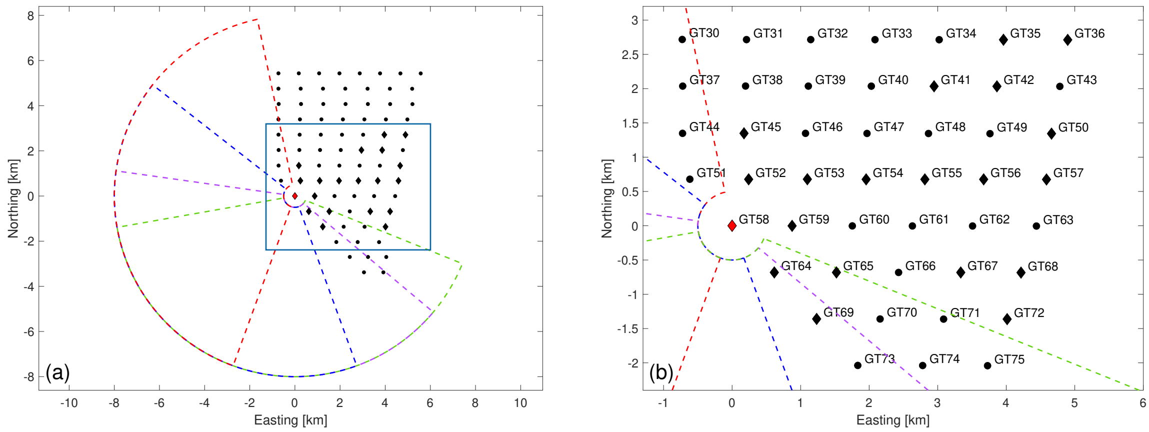

Figure 1: (a) Layout of the wind farm Global Tech I with turbine positions visualized as black dots. The lidar position is visualized as red diamond and scanned sectors as coloured dashed lines. The Cartesian grid is centred around the lidar's position. The blue rectangle indicates the zoomed-in region shown in (b). The 20 turbines with the highest forecast availability are marked as black diamonds.

vectors and weighted equally. Wind speed forecasts were transformed to power forecast using wind turbine power curves. Power forecasts were calibrated using an EMOS (ensemble model output statistics) approach [12]. We will refer to this approach as observer-based forecast in the following. Both the data set and the forecasting methodology are described in more detail in [6] and [10].

2.2. Hybrid methods combining the observer-based forecast and persistence

As the atmospheric conditions were found to have a strong impact on the skill of the LF and OF and their performance in comparison to persistence [6, 10], we developed a hybrid methodology that aims to consider these different conditions. We used the Obukhov length L, the wind speed at 100 m height $u_{100\,\mathrm{m}}$ and the wind direction at 100 m height $\chi_{100\,\mathrm{m}}$ extracted with 10 minute resolution from a WRF (Weather Research and Forecasting Model) simulation at the location of the offshore wind farm Global Tech I [13]. The wind speed was transformed into turbine power P utilizing power curves constructed from high-frequency SCADA data. L, $\chi_{100\,\mathrm{m}}$ and P were further linearly interpolated to the forecast's time steps and used to categorize the forecasts into power and stability bins. Hereby, we distinguished between unstable ($-1000\,\mathrm{m} < L < 0\,\mathrm{m}$), neutral ($|L| >= 1000\,\mathrm{m}$) and stable ($0\,\mathrm{m} < L < 1000\,\mathrm{m}$) stratification and low ($0.05\,P_\mathrm{r} \leq P < 0.2P_\mathrm{r}$), medium ($0.2\,P_\mathrm{r} \leq P < 0.9\,P_\mathrm{r}$) and high power production ($P \geq 0.9\,P_\mathrm{r}$). We further considered if turbines were placed inside the wake of surrounding turbines or not. This information was extracted from $\chi_{100\,\mathrm{m}}$ and the wind farm layout. As we expect wake influences to be less important for the upper and lower power ranges and we wanted to avoid overfitting as a consequence of too little available data, we made this distinction only for the middle power range.

Subsequently, each bin was bisected into a training data set, that was used to optimize the hybrid forecast, and a test set to evaluate the methodology. To combine the OF and persistence we used two approaches. For the first approach, referred to as binary method in the following, the root-mean-squared error (rmse) of each bin's training data set was compared for the OF and persistence. The method with the lower score was deemed superior for the selected bin and applied to corresponding situations in the test data set. For the second approach, referred to as weighting method, the observer-based forecast $\mathrm{fc}_{\mathrm{OF}}$ and persistence $\mathrm{fc}_{\mathrm{pers}}$ were weighted and averaged within each bin b. The assigned weight w_b was hereby determined using a cost-function J_b

$$J_b = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathrm{fc}_{\mathrm{weighting},i} - \mathrm{obs}_i)^2} \tag{1}$$

that minimizes the rmse of the training data set with

$$\mathrm{fc}_{\mathrm{weighting},i} = w_b\,\mathrm{fc}_{\mathrm{pers},i} + (1 - w_b)\,\mathrm{fc}_{\mathrm{OF},i}, \tag{2}$$

the number of considered time steps N and the observation obs.

Hybrid forecasts were evaluated in terms of rmse, bias and rmse skill score (ss)

$$\mathrm{rmse\ ss} = 100\left(1 - \frac{\mathrm{rmse}}{\mathrm{rmse}_{\mathrm{ref}}}\right). \tag{3}$$

The rmse of the reference forecast is denoted $\mathrm{rmse}_{\mathrm{ref}}$. For a perfect forecast with rmse $= 0$ or rmse $\ll \mathrm{rmse}_{\mathrm{ref}}$ the skill score equals 100 % and is negative for rmse $> \mathrm{rmse}_{\mathrm{ref}}$.

For the analysis we focused on turbines GT30-GT75 shown in Figure 1 (b) as the forecast availability of the remaining turbines, positioned in the northerly and southerly region of the wind farm, is low due to insufficient lidar and SCADA coverage.

2.3. Extension to a one year-long data set

To assess the usefulness of the OF, the hybrid methods and persistence, they have to be evaluated over a longer period of time. A meaningful time period would, for instance, be one year. Hereby, the atmospheric conditions need to be considered as they strongly impact the methods' forecast skill. The duration of the measurement campaign and therefore the available forecasts do not cover a sufficiently large time period. We therefore extrapolated the available data set to a whole year mimicking typical atmospheric conditions at the wind farm site. To do so, we utilized wind speed, wind direction and atmospheric stability information at GT I extracted from WRF in the period 2010-2019 [13]. This information was binned using stability regimes as described in Section 2.2, the wind speed intervals $[4; 8; 12; 35]\,\mathrm{m\,s^{-1}}$, and the wind direction intervals $[170; 200; 230; 260; 290; 320; 350]^\circ$. Wind directions in the range from 350° to 170° were not considered. This is based on the assumption that a single lidar device is able to sufficiently cover an azimuth range of approximately 180°. In accordance with that, the available forecasts are not able to sufficiently cover a wider set of wind directions. Also the available forecasts were binned according to these criteria and using data extracted from WRF. The number of average occurrences per year was then determined for each bin and the corresponding number of forecasts was drawn from the binned forecasts using a random sampling technique with replacement. A bootstrapping method was applied to determine the significance of the results. In cases with no forecasts available within a certain bin, we first enlarged the wind direction bin by allowing a selection from the two adjacent bins. In a second step, we enlarged the wind speed bin in the same manner, and finally only considered stability information. This procedure was applied as the highest impact on forecast skill is expected to be related to atmospheric stability based on results from previous work [6]. Extended forecasts were evaluated in terms of rmse and bias.

3. Results

3.1. Hybrid methods combining the observer-based forecast and persistence

The OF, persistence and the two introduced hybrid methods were evaluated using 1-minute-mean SCADA data. Hereby, only situations under normal operating conditions were considered.

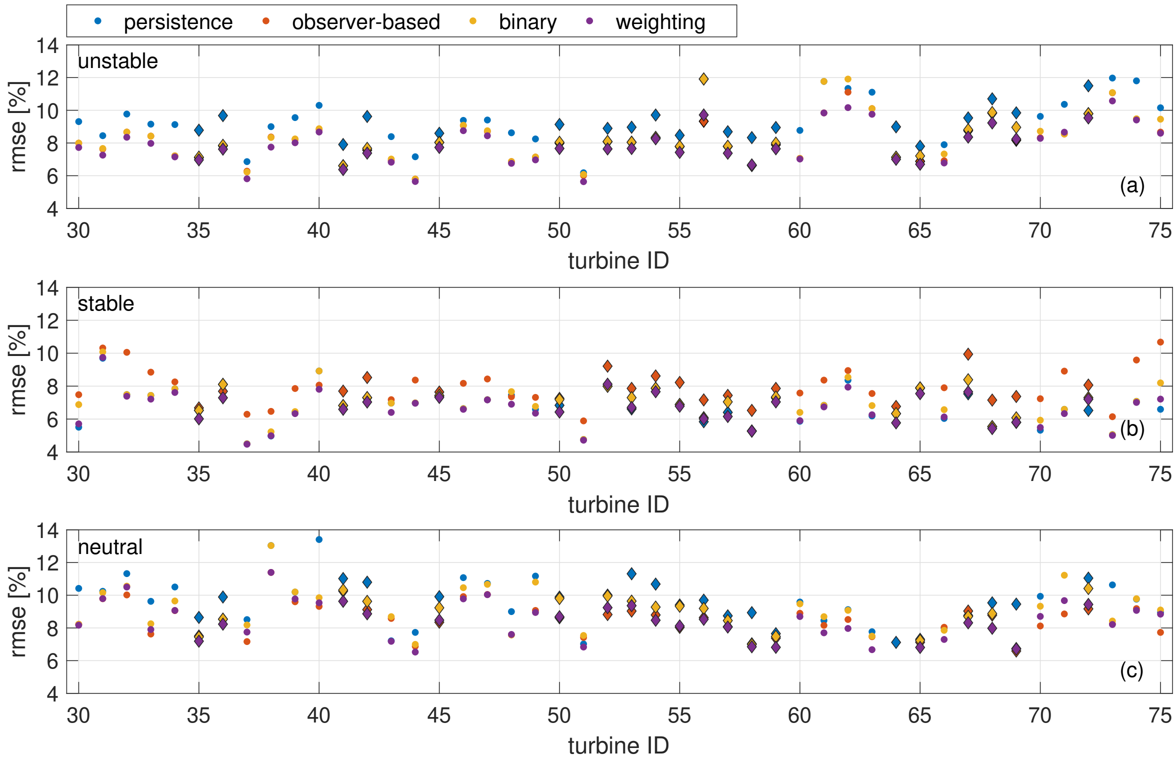

Figure 2: The rmse of persistence, the OF and the hybrid methods for the turbines GT30-GT75 of GT I distinguishing between (a) unstable, (b) stable and (c) neutral atmospheric conditions. The rmse is given in % of the wind turbines' rated power. The 20 turbines with the highest forecast availability are marked by ♢.

Figure 2 compares the rmse of all forecasts for turbines GT30-GT75 for (a) unstable, (b) stable and (c) neutral atmospheric conditions. The number of valid forecasts varies for the different turbines as a consequence of wind speed, wind direction, wind farm layout and the lidar trajectories. In unstable cases persistence has the largest rmse for most of the analysed turbines. The weighting method shows an improved skill compared to the OF, while the binary method shows scores similar to the OF, outperforming it in few cases. Similar as observed in previous work persistence outperforms the OF during stable stratification [6, 10]. For most turbines, the skill of the binary approach is lower than that of persistence but higher than that of the OF. The weighting approach performs only slightly better and is able to outperform persistence for some of the turbines. In neutral situations generally the weighting method performs best, followed by the OF and binary method.

After comparing the scores of all methods for the 46 selected turbines of GT I we will focus on the 20 turbines with the highest forecast availability in the following. These turbines are marked as ♢ in Figure 1 (b) and Figure 2.

In order to analyse how the two hybrid approaches select respectively weight the individual methods under different conditions, we depict the distribution of forecasts per bin, i. e. the number of forecasts available in the training data set, as a bar plot for the different stability classes and power regimes in Figure 3 (a). The median value is shown as horizontal line and the boxes include the 25% to 75% quantiles. The dashed lines extend to the minimum and maximum values. Moreover, in Figure 3 (b) the results of the selection algorithm for both the hybrid models are depicted. The yellow markers indicate the share of the analysed turbines for

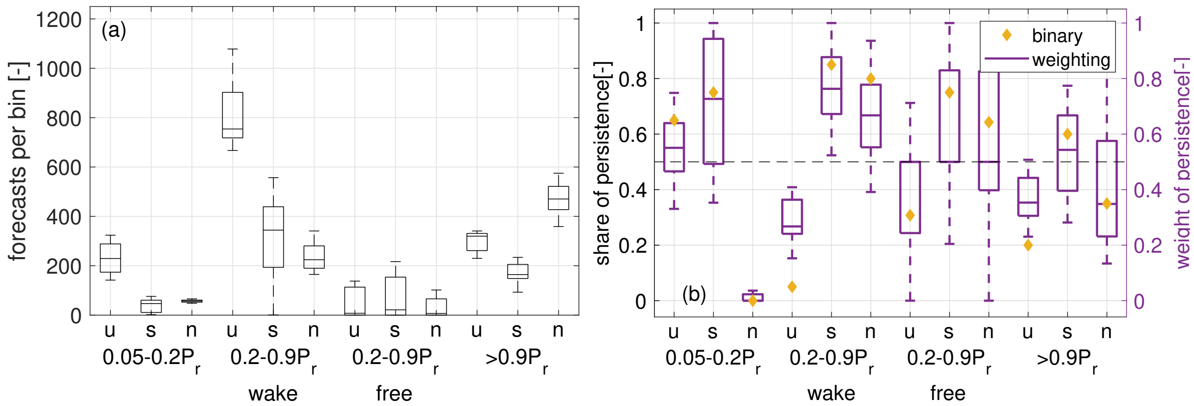

Figure 3: (a) Number of valid forecasts for different stability (u,s,n) and power bins. Horizontal lines show the median, the boxes include 50 % of the data. Dashed lines extend to the maximal and minimal values. (b) The share of turbines for which persistence was selected as superior forecasting method in the binary method for the different bins is shown in yellow. The weight of persistence in the weighting approach is visualized as box plot. In both subfigures only the 20 turbines with the most available forecasts are considered.

which persistence was deemed superior for the binary hybrid method and thus selected for the respective stability and power bin. The boxes represent the weight put onto persistence for the weighting approach using Equations (1) and (2). Overall, the results of the two methods agree well. The OF was weighted stronger and accordingly selected more often for unstable cases and in particular for the medium power bin and waked turbines. Low power cases are an exception, here persistence was weighted stronger than the OF. During stable cases persistence dominates the hybrid methods. For high power regimes, the weights determined for the weighted approach are distributed more closely around 0.5 than for other power regimes. Weights for free-flow and wake-influenced wind turbines are similar, with a larger spread for free-flow cases. This is probably a consequence of lower data availability, which is related to the fact that many of the considered turbines are placed within the wind farm and likely always impacted by wakes.

To answer the question to what extent and under which conditions the observer-based forecast and hybrid models are able to outperform persistence, Figure 4 visualizes the average rmse skill score (ss) with respect to the rmse of persistence and the median bias of the selected turbines for different forecasts and bins. The 50 % confidence intervals are depicted as error bars. For several bins, mainly stable atmospheric conditions, the rmse skill score of the OF is strongly negative. The rmse ss of the binary method often lies close to 0. In these cases persistence was selected for most of the turbines, and thus only little improvement or decline can be observed. Negative skill scores in these cases indicate that the selection of the OF over persistence does not improve forecasts in the test data set. While the skill score of the binary method can, in the best case scenario, equal the positive ones of the OF, the weighting method has the option to outperform the OF skill score by combining the two forecasts. It outperforms persistence for most bins, the largest improvements are observed for neutral and unstable cases. Here, the OF often already has a positive rmse ss, thus outperforms persistence. Also for stable cases in the medium, wake-influenced power regime and the high power regime the rmse ss of the weighting method is positive. Even though the OF's skill is worse than that of persistence for these cases, it can add some value to a hybrid method. In the medium, wake-influenced power regime the OF's and persistence's median bias have opposite signs. This results in a lower absolute median bias of the hybrid methods, in particular of the weighting approach. For most other bins, systematic errors of the hybrid methods are above those of either the OF or persistence. For several bins, the confidence intervals are very broad, questioning the significance of the results. This is most distinct for free-flow situations and likely related to low data availability (cf. Figure 3 (a)).

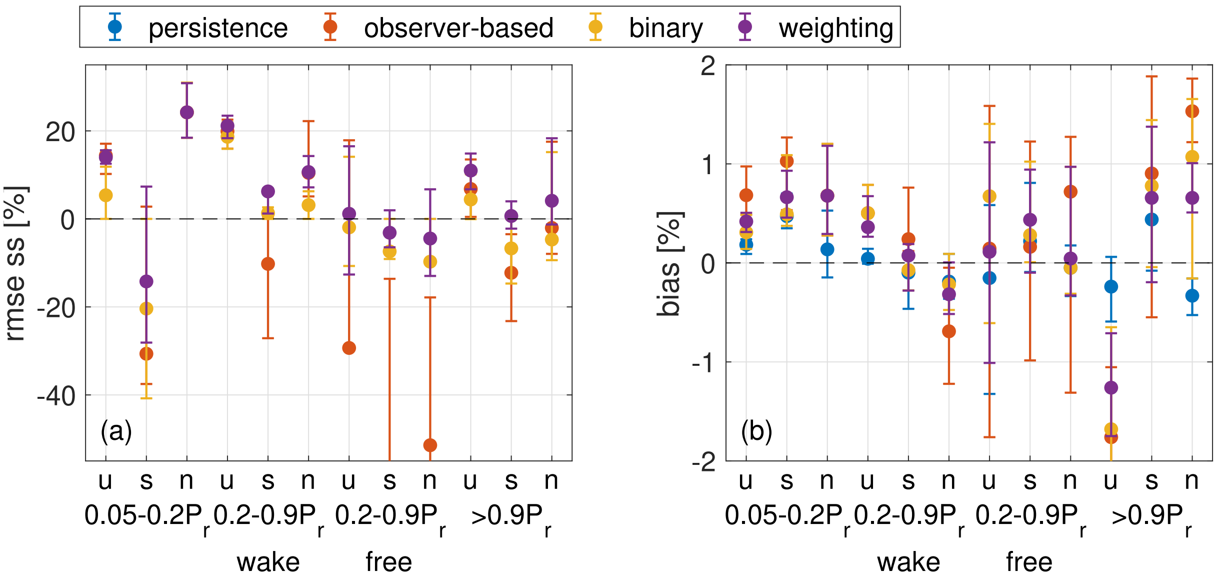

Figure 4: Median rmse skill score (a) and bias (b) in % of the turbine's rated power for different stability and power bins and forecasting approaches. Error bars depict 50% confidence intervals. In both subfigures only the 20 turbines with the most available forecasts are considered.

3.2. Extension to one year-long data set

As a basis for the extension to a one year-long data set, in Figure 5 the typical atmospheric conditions extracted from WRF simulations at the location of GT I for the years 2010-2019 are summarized. The main wind direction was identified as west/south-west as visualized in the wind rose in Figure 5 (a). The considered wind directions from $170° - 350°$ cover around 65.6 % of the whole data set. Further, a stability wind rose is shown in Figure 5 (b). The share of stable situations is largest for westerly winds and lowest for north-easterly wind directions.

In Figure 6 we show the stability distribution, wind speed and wind direction distribution of an average year at the location of GT I as extracted from WRF (leftmost). The distributions presented here are the binning results described in Section 2.3. To evaluate how accurate the atmospheric conditions were mimicked using the resampling technique and available forecasts, we additionally show distributions of those turbines of GT I highlighted in Figure 1 (b) and Figure 2 (◇). Differing wind speed and wind direction values and in particular a varying number of available forecasts for individual turbines can cause different distributions of wind speed, wind direction and atmospheric stability for the different turbines. Further, the bins shown in Figure 6 are finer than those applied for resampling (cf. Section 2.3), causing a difference between WRF and wind turbine distributions. In general, the atmospheric conditions correspond well to the ones extracted from WRF. Very stable and very unstable cases are slightly under-represented in the artificial data set. Similarly, the number of wind speeds from $4\,\mathrm{m\,s^{-1}}$ to $6\,\mathrm{m\,s^{-1}}$ is too low. Wind directions from $170° - 230°$ are not represented well due to limited forecast availability for these wind directions. As we use neighbouring wind speed bins as an alternative, wind directions $230° - 260°$ are over-represented.

After confirming the adequate representation of typical atmospheric conditions in the artificial data set we evaluate the forecast skill. In Figure 7 we show the rmse of the wind turbines GT30-

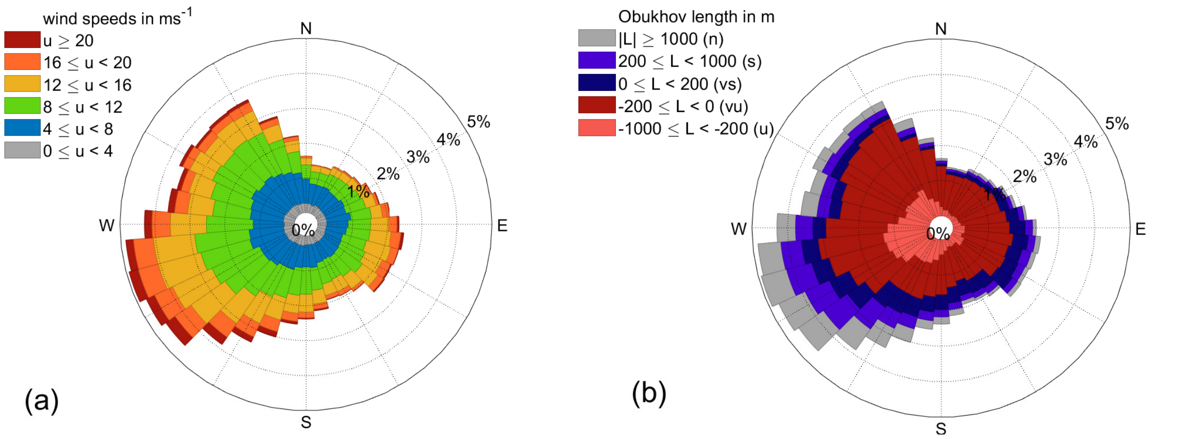

Figure 5: (a) Wind rose extracted from WRF at the location of GT I and a height of 100 m for a 10-year period. (b) Stability wind rose extracted from WRF distinguishing between very stable (vs), stable (s), neutral (n), unstable (u) and very unstable (vu) cases.

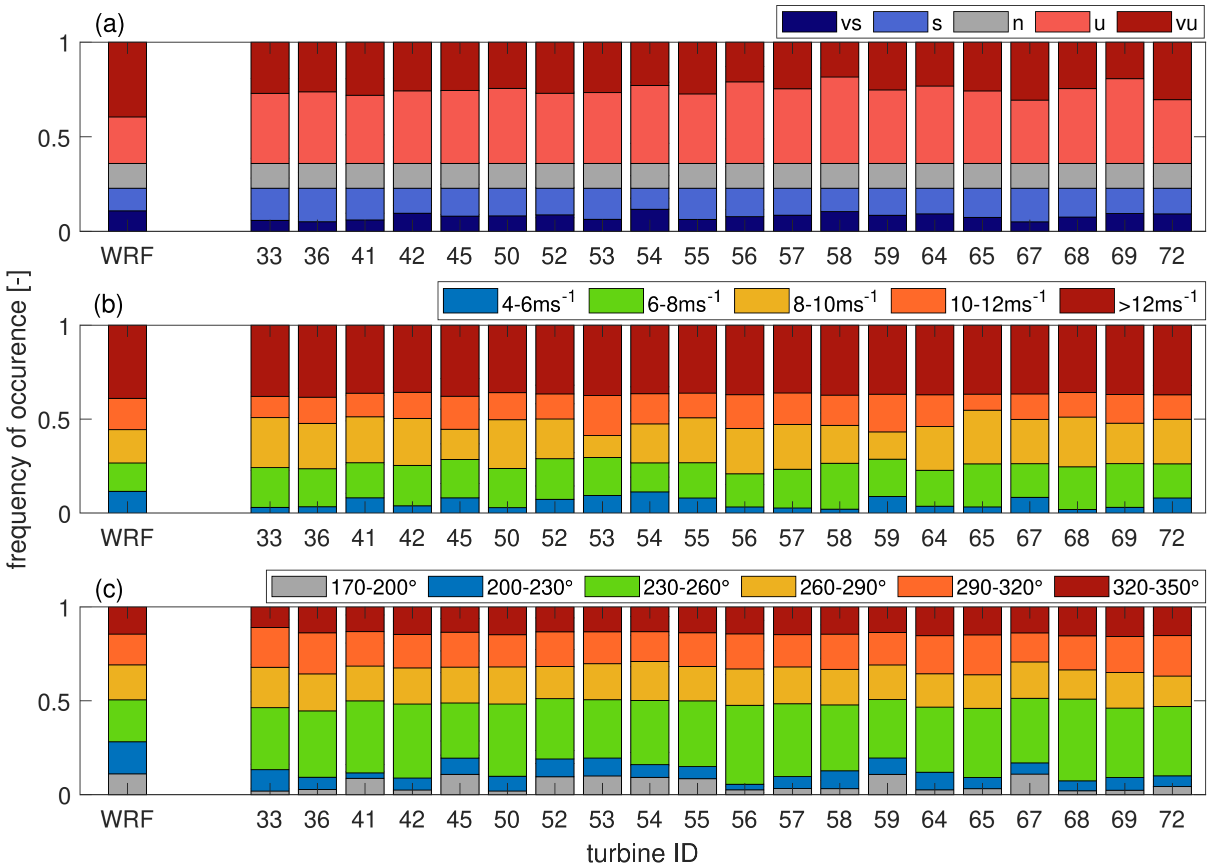

Figure 6: (a) Atmospheric stability, (b) wind speed and (c) wind direction distributions extracted from WRF (lefmost) and of the extended forecasts of the 20 turbines highlighted in Figure 2 and Figure 7.

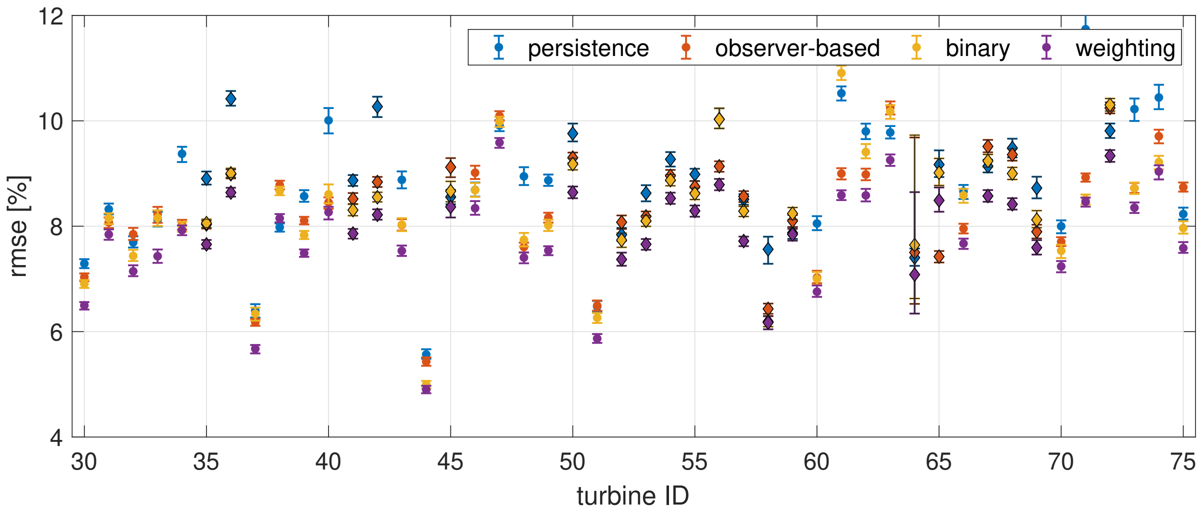

Figure 7: The rmse of the extended versions of persistence, the OF and the hybrid methods. The rmse is given in % of the wind turbines' rated power. Markers represent the 50 % quantile and the 95 % confidence intervals are visualized as error bars. The 20 turbines with the highest forecast availability are marked by ♢.

GT75 in % of their rated power for the different forecasting approaches and the extended, artificially created data set. Markers represent the 50 % quantile and the 95 % confidence intervals are visualized as error bars. Moreover, average rmse values are summarized in Table 1. Again, the turbines with the highest number of valid forecasts are marked as ♢. For 44 of the 46 turbines the weighting approach provides the most skillful forecast. Considering only the turbines with the highest availability this holds for 19 out of 20. Also the average rmse of both analysed turbine subsets is lowest for the weighting approach. For many turbines persistence can be outperformed significantly by all methods. The confidence intervals are generally very narrow and results can therefore be interpreted as significant. Figure 6 shows that the share of unstable cases of the considered data set lies at around 64.1 %. As shown in Section 3.1, the OF and hybrid methods are able to outperform persistence during those cases.

Table 1: Average rmse of persistence, the observer-based forecast and the two hybrid methods for the artificially created data set in % of rated power. Lowest scores are shown in bold.

	persistence	**observer-based**	**binary**	**weighting**
GT30-GT75	8.82	8.34	8.33	**7.83**
high availability, 20 turbines	8.96	8.55	8.56	**8.06**

4. Discussion

4.1. Hybrid methods combining the observer-based forecast and persistence

The two hybrid methods introduced in this work rely on a training data set that is divided into several bins characterizing the local atmospheric conditions. Due to a limited amount of available forecasts, we were restricted to a relatively small number of bins in this study. For future work, a finer distinction between atmospheric conditions might be beneficial. However, in this case study it would have likely led to overfitting. Also, larger training data sets might have improved the results. Generally, it should be further analysed what size of training data set is sufficient and whether weights for the weighting approach deduced here could be used to generate forecasts over longer time periods. We expect this to be less critical for the binary approach.

The weighting method has advantages over the binary method as it can consider both forecasting methodologies. This allows to balance systematic errors of the individual forecasts (cf. Figure 4 (b)). The binary approach is based on a binary selection system. If the choice based on the training data set is wrong for the test data set, the forecast skill is maximally degraded. For the weighting approach, this might also happen if one forecast is assigned a weight of 0. However, it is more likely that results will benefit from both methods. That means using the binary method is associated with a higher risk of reducing forecast skill, while its possibility to reduce errors is limited. Overall, we found atmospheric stability to have the largest impact on the forecasts' weighting. Due to limited data availability we could not draw significant conclusions regarding the weighting for free-flow and wake-influenced turbines. Results indicate that a higher weight of OFs compared to persistence during wake-influenced cases is more skillful. In agreement with previous results, persistence is more skillful during stable atmospheric conditions and thus assigned a higher weight [6]. It stands out that for low power cases and unstable conditions persistence is also weighted more compared to the OF. This might be related to lower absolute changes in wind speed observed during these situations, possibly leading to an advantage for persistence. In [6] persistence outperformed a lidar-based forecast for 5-minute wind speed increments smaller than $0.5\,\mathrm{m\,s^{-1}}$. For high power regimes ($> 0.9P_r$) the difference in weighting between OF and persistence is less clear than for medium and low power values. Here, many situations with wind speed above rated wind speed were considered. For those cases wind speed errors have only little impact on the skill of power forecasts and the skill of both forecasting methods converges.

The complementary use of persistence and an OF was found to be most beneficial during unstable atmospheric conditions. Even during stable atmospheric conditions the OF can add additional skill to persistence.

4.2. Extension to one year-long data set

The extension of the forecast to a longer data set is based on the prerequisite that enough forecasts are available within each bin. The limited amount of data in this analysis restricted the bins that could be used and resulted in slightly different conditions for the artificial forecasts compared to the WRF data. The wind direction mainly influences the impact of wakes on individual turbines. The results indicate that in particular wake-influenced situations might benefit from lidar and SCADA data. We thus need to assume that an inaccurate representation of the wind direction distribution might to some extent influence the results. Also the power regime was found to influence forecast skill. We suspect that during higher wind speeds mainly the observer-based forecast and hybrid methods gain benefits. However, as we were able to reconstruct atmospheric stability distributions well and stratification was found to have the strongest impact on forecast skill, we consider our results meaningful despite these limitations. The narrow confidence intervals support this notion (cf. Figure 7). The results achieved by the extension to an artificial but representative typical year prove the value of the weighting approach for an operational use.

The atmospheric conditions observed at the wind farm GT I are not untypical for offshore locations. We, therefore, assume the conclusions drawn in this study are in general applicable also to other offshore sites in the North Sea. To draw more detailed conclusions regarding the value of observer-based forecasts as a supplement to persistence one needs to analyse the conditions at each site individually.

5. Conclusions

This study implemented two hybrid methods, combining an observer-based forecast and the statistical method persistence, aiming to understand the benefits of their complementary use. The binary method used a binary selection of either of the two forecasts and the weighting method applied a weighting function. Generally, the weighting approach achieved more skillful results. It is more flexible and can account for systematic errors by combining the individual forecasts. The forecast skill of both the individual and hybrid methods was mainly influenced by atmospheric stability. The highest forecast skill compared to persistence was observed during unstable atmospheric conditions. However, also during stable conditions the observer-based forecast can add value to persistence. Despite the limited forecast availability, atmospheric conditions at the offshore site could be replicated accurately using the resampling technique. As the wind farm site is dominated by unstable stratification, the observer-based and hybrid methods can also significantly increase forecast skill when evaluated over longer time periods. Highest forecast skill was achieved by the weighting approach for both analysed turbine subsets, which proves its value for an operational use.

Acknowledgments

We thank the German Federal Environmental Foundation (DBU) as this project received funding within the scope of their PhD scholarship program (Grant Nr. 20018/582). The lidar measurements and parts of the work were performed within the research projects OWP Control (Ref. Nr. 0324131A), WIMS-Cluster (Ref. Nr. 0324005) and 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 Global Tech I Offshore Wind GmbH for providing SCADA data and access to their wind farm and thank them for supporting our work. We acknowledge Met Office for making the OSTIA data set available. We thank Stephan Stone for conducting the measurement campaign and Andreas Rott for supporting the development of the observer-based forecasting methodology.

References

[1] Würth I, Valldecabres L, Simon E, Möhrlen C, Uzunoğlu B, Gilbert C, Giebel G, Schlipf D and Kaifel A 2019 *Energies* **12** 712
[2] Gneiting T, Balabdaoui F and Raftery A E 2007 *Journal of the Royal Statistical Society* **69** 243–268
[3] Chang W 2014 *Journal of Power and Energy Engineering* 161–168
[4] Valldecabres L, von Bremen L and Kühn M 2020 *Wind Energy* **23** 1–23
[5] Theuer F, van Dooren M F, von Bremen L and Kühn M 2020 *Wind Energy Science* **5** 1449–1468
[6] Theuer F, van Dooren M F, von Bremen L and Kühn M 2021 *Meteorologische Zeitschrift* **31** 13–29
[7] Pichault M, Vincent C, Skidmore G and Monty J 2021 *Energies* **14** ISSN 1996-1073
[8] Valldecabres L, Nygaard N, Vera-Tudela L, von Bremen L and Kühn M 2018 *Remote Sensing* **10** 1701
[9] Rott A, Petrović V and Kühn M 2020 *Journal of Physics: Conference Series* **1618** 062067
[10] Theuer F, Rott A, Schneemann J, von Bremen L and Kühn M 2022 *Wind Energy Science, submitted*
[11] Werner C 2005 *Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere* (New York, NY: Springer New York) chap 12 - Doppler Wind Lidar, pp 325–354 ISBN 978-0-387-25101-1
[12] Thorarinsdottir T L and Gneiting T 2010 *Journal of the Royal Statistical Society* **173** 371–388
[13] Dörenkämper M, Olsen B T, Witha B, Hahmann A N, Davis N N, Barcons J, Ezber Y, García-Bustamante E, González-Rouco J F, Navarro J, Sastre-Marugán M, Sīle T, Trei W, Žagar M, Badger J, Gottschall J, Sanz Rodrigo J and Mann J 2020 *Geoscientific Model Development* **13** 5079–5102