Review: Theuer et al.: Enhancing minute-scale lidar-based power forecasts of offshore wind farms towards an operational use

Summary:

The paper extends previously established methodology for lidar-based forecasts (LF) for offshore wind energy and focuses on extended forecast horizons of up to 30 minutes. The work investigates and compares different methods for height corrections and establishes criteria to identify conditions under which the LF can be successfully applied. While several results generally confirm findings previously reported in the literature, the detailed analysis of factors influencing the forecast skill, such as wind vector age/propagation duration, and the consideration of long forecast horizons add novelty and make it worth publishing. However, several points should be addressed before publication:

- One of the key novelties of the paper is the extension of the forecast horizon up to 30 minutes. The authors also emphasize this aspect in the abstract and the introduction. The results are somewhat sobering for the longest forecast horizons and no positive skill score could be achieved. While this might be disappointing for applications, it is an important finding. However, the skill scores for very long forecast horizons are not mentioned in the abstract nor in the conclusion. This creates an inconsistency in the overall presentation. I suggest placing this more prominently in both sections and clearly stating the maximum forecast horizons for which a meaningful skill score can be achieved.

- The authors identify the farm-level forecast as the most important for applications. However, the presentation of the results for the whole wind farm is rather brief. I suggest expanding this section and discussing the presented results in more detail. If possible, add results on, e.g., the effect of the height correction on the forecast for the entire wind farm.

- In some places, descriptions of the methodology and conclusions drawn from the results lack precision.

- In some places, results and conclusions are presented as statements and omit the presentation of (quantitative) evidence.

- Towards the end of the results section and in some parts of the conclusion, the text lacks clarity and is hard to understand.
- In some places, English language editing could be improved. I did not include this in the review.

These points are also included in the specific comments below (except English editing).

Specific comments:

- L43: The authors state: "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 processing time of a few minutes for forecast generation and application is required." This is the central motivation for why the extension of the forecast horizon is relevant. However, no further explanation and/or literature is given. Please add a short explanation, ideally including literature, to support your statement.
- L82: "To avoid a frequent change of trajectories, the wind direction intervals were defined to overlap in a 10° sector." Is this intended as a hysteresis? Please rephrase for clarity and name it if so.
- L81: "These sectors were optimized for lidar-based forecasting considering the wind direction, the lidar position and the wind farm layout." An optimization suggests an objective quantitative procedure. Please provide a quantitative description of the optimization process (objective, constraints, method). If no objective and quantitative approach was used, please rephrase.
- L80: "During the campaign, it was alternated between five horizontal PPIs and a set of profile-PPIs." What is "it"? The scan pattern? Please clarify.
- Table 2: Splitting this table into two tables (one for horizontal PPI and one for profile PPI) would increase readability.
- Figure 2: There are no explicit panel labels (a) and (b) in the figure (only left and right). Please ensure the figure panels are labeled to match the caption and provide the projection/coordinate system for the left panel.
- L127: "In this work, k_{tend} was optimized in a sensitivity study and defined as..." Please provide sufficient details so the sensitivity study is reproducible. What was the optimization criterion/cost function, parameter search space?
- L132: How were the inclination data processed/averaged? The inclination sensor measures acceleration relative to the gravitational field (I

assume). High-frequency vibration-related accelerations from the turbine could contaminate short-interval inclination estimates and introduce "phantom-inclinations". While distinguishing these effects may be difficult, this potential issue should at least be discussed.

- L169: "While in this work we analysed deterministic forecasts only..." I suggest moving this to the discussion rather than methods/results, as probabilistic methodology is not applied here.
- Figure 3: If the aim is to illustrate scatter, please state this explicitly. It would also help to use a colour bar to indicate point density.
- L181: I am not sure if "deterministic" is the right word here. Maybe use "average" instead.
- L253: "...recognising that the forecast characteristics for neutral stratification, e.g. classified by $|L| > 500$ m, could differ." Why did you not include a separate neutral class if you recognise the results could differ?
- L253: "With increasing lead time, the forecast availability significantly reduces." This is important for applications and could be displayed and discussed in more detail. Percentages of forecast availability over the investigated period would be easier to interpret than absolute counts.
- Figure 4: It is very hard to distinguish the markers. Please use more easily distinguishable symbols/line styles.
- L279: "Comparing the rmse and rmse ss of individual turbines, it can be noted that free-stream turbines for main wind 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 this may have been observed, it is not visible in the presented results. Either modify the figures to illustrate this or provide additional quantitative results supporting these statements.
- L283: You attribute the observed biases to the limited lidar range. Could increased pointing uncertainty, measurement height errors and wind-profile errors also play a role? If not, please discuss why not.
- L336: "...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." Please define and explain "scan ID" more clearly and how it is assigned.
- L363: "Additionally, the number of valid forecasts..." This sentence is cumbersome and hard to read, and it reads like a table header. Please rephrase and explain the quantities while discussing the results separately.

- L366: “The skill at a 15-minute horizon holds significant value when aiming at the application of the forecast for trading and system operation.” This appears out of context here. A short explanation of why 15 minutes is critical would help. Also, many other forecast horizons are discussed—what is their intended use?
- Figure 12: Introduce a symbol for propagation time/duration and add it to the legend. You use ΔT_{prop} in Table 3 but do not introduce it in the text before.
- End of Section 3.4: I think the farm-level results are the most relevant for applications. Nevertheless, very little space is given to explaining the results, and they are cramped between long case definitions. I suggest allocating more space to explain the results. At present, the last two paragraphs of Section 3.4 are quite hard to read.
- Table 3: Much of this information is already displayed in Figure 12. Consider condensing the table or moving some content to the figure/caption.
- Section 3.4: Use “maximum” instead of “maximal.”
- L396: “In those cases, the forecast uncertainty is not dominated by the height correction...” This implies that height correction dominates at shorter lead times. Consider rephrasing to “In those cases, height correction is less important compared to other factors.”
- L416: “The estimation of the measurement height using the tilt model was similarly accurate to using the inclination sensor data.” The measurement height itself was not evaluated in this work. Therefore, this conclusion cannot be drawn as stated. You showed that wind speed extrapolation led to similar improvements in forecast skill for both methods.
- L444–L450: “If only few scans contribute to a forecast...” The point is hard to follow. Please rephrase for clarity. Also check whether “enhanced” is the right verb in some sentences.
- L502: “Our results also confirmed the dependence of forecast skill on turbine position... Free-stream turbines can be predicted more accurately than inner-farm ones.” This is stated but not shown in the displayed results (see comment above). Please provide supporting evidence or soften the claim.
- L511: “As ill-predicted ramp events can become very costly both in electricity trading...” You argue economic value via ramp events; however, you did not investigate them here. It would be better to explain how the KPIs investigated in your study contribute to the economic viability of LFs, or clearly point to planned future work.
- L516: “...of a minute-scale wind speed and power forecast.” Add “lidar-based” (or similar) for clarity.

- Conclusion: Why do you not mention farm-level forecasts here? You state yourself that they are most relevant for applications.
- L528: "Considering these forecast characteristics has proven useful for the assessment of wind farm power forecasts and can be beneficial to end-users." This sentence seems too vague. Consider specifying how it benefits end-users (e.g., selection rules, extended horizon where persistence is outperformed).