Reviewer 1

The current version of the manuscript "Enhancing turbulent fluctuation measurement with tailored wind lidar profilers" has been fully revised and is substantially different from the version initially submitted. The most significant change is that now only one prototype lidar with faster sampling rate is being tested against the commercial version of the WindCube 2.1 and that a great effort has been put into quantifying the effect of noise on the measurements. The narrowed scope appears reasonable and has the potential to result in a better paper.

A strength of the manuscript is the work put into a quantification of noise in the LOS data from the WindCube. Though, it is unclear why the authors did not write about the CNR value that is provided by the instrument as a standard parameter and could have been used to compare the noise levels of the prototype lidar and the commercial version.

The main weakness of the manuscript lies in the assumption that a reduction in sampling frequency leads to a reduction in variance of the measurement data, which is not true. This false assumption that increasing the sampling rate could capture additional energy associated with smaller eddies leads the interpretation of the experimental data into a wrong direction. Instead, more focus should be put on the relationship of intra beam and temporal averaging and how it is influenced by the prevailing mean wind speeds.

In its current form, the manuscript is not ready for being accepted by WES and it should be reconsidered after major revisions. Please note that the following comments are not capturing all aspects that should be improved and that a revision should be done with care before submission.

We sincerely thank the reviewer for its detailed and constructive feedback. We appreciate your recognition of the improvements made in narrowing the scope and quantifying the noise in the LOS data. Your comments have been very helpful in identifying areas that required clarification, particularly regarding the use of CNR, the assumptions around sampling frequency and variance, and the role of intra-beam and temporal averaging. We have addressed these points carefully in the revised manuscript and made substantial changes to improve both clarity and scientific accuracy.

Specific comments:

Response to the reviewer:

It is good practice to acknowledge the referee's effort put into reviewing the manuscript. The authors missed this opportunity which is discouraging. Further, the response to the reviewer is suffering from mistakes, e.g., "We are right." instead of "You are right." and statements that are not covered in the updated manuscript, e.g., "We have implemented logarithmically spaced [sic] frequency bins [...] See Fig. 7b.".

We sincerely appreciate the time and effort you dedicated to reviewing our manuscript. We deeply regret that we did not explicitly acknowledge your contribution in our initial response, and we apologize for this oversight. Your feedback is very valuable to us, and we are grateful for the constructive comments you provided.

We also apologize for the errors in our response, particularly the phrasing of "We are right" instead of "You are right." This was a mistake on our part, and we will ensure that our communication is more respectful and accurate in the future. Furthermore, we recognize that we incorrectly referenced changes that were not fully reflected in the manuscript, such as the claim about the logarithmically spaced frequency bins in Figure 7b. We have corrected these inconsistencies and updated the manuscript accordingly.

Thank you again for your thorough review and for pointing out these issues. We hope that the revised manuscript meets your expectations.

1.: The introduction gives some valuable insights into the history behind the topic, but the state of the art is insufficiently covered. Please add the most relevant and significant findings from the cited literature instead of just listing it in groups. The section should end with a guidance through the structure of the paper.

We have substantially revised the Introduction to better reflect the state of the art. Specifically, we have now included a concise summary of the most relevant and significant findings from the cited literature, rather than listing them in groups (Lines 79-89, page 4). Additionally, as recommended, we have added a brief paragraph at the end of the Introduction to guide the reader through the structure of the paper (Lines 96-104, page 4).

2.1: Include a table with a comparison of the two lidar configurations showing parameters like sampling rate, accumulation time per LOS, number of samples per 30 min, range gate...

We have added Table 1 (page 6) to the manuscript, which provides a side-by-side comparison of the two lidar configurations. The table includes key parameters such as the sampling rate, accumulation time per LOS, number of samples per 30 minutes, and probe length.

2.1: The authors should reflect on the relationship between the industry demand for TI data (10 min.) and the variance of the u-component of the wind (30 min.) provided by the methods described in the paper.

We have added this in the Introduction: "This enhancement is assessed for its impact on measuring mean wind speed, data availability, and along-wind variance and its square root, i.e., the standard deviation. The latter is particularly important, as it is used in the wind power industry to compute turbulence intensity (TI), a critical metric for turbine load assessment, site suitability, and energy yield predictions." (Lines 91-94, page 4), and this:

“The selection of a 30-min window, rather than the standard 10-minute interval commonly used in the wind energy industry, was guided by the aim of reducing random errors in turbulence measurements, following the recommendations of Lenschow et al. (1994).” (Lines 165-167, page 8).

2.2.1: This subsubsection is the only content of subsection 2.2. This does not make sense.

As suggested, we have removed subsubsection 2.2.1, since it was the only content within subsection 2.2, making the subdivision unnecessary. The content now appears directly under subsection 2.2 for improved clarity and structure.

2.2.1: There are wind turbines only 210 m away from the lidars, so there is no "undisturbed winds from almost all sectors". Please explain if only wind from the wind turbine's upstream direction was used in the study.

We have added a wind rose to illustrate the sector contaminated by the turbine wake, which is highlighted by the blue shaded areas in Fig. 4 (page 8). The wind sectors selected for the present analysis, indicated by the gray shaded areas in the same figure, were carefully chosen to lie outside the contaminated region.

2.2.1: Please describe the purpose of creating smaller subsets of data sampled at 0.25Hz and 1 Hz respectively. If the sonic was configured with higher sampling rate, the entire dataset could be used with 0.25Hz and 1Hz. This is unclear.

To clarify, only the sonic anemometer data were resampled to 1 Hz in order to match the sampling rate of the prototype lidar. The datasets from the commercial and prototype lidars themselves were not resampled; their native sampling rates of 0.25 Hz and 1 Hz, respectively, were retained. Consequently, each 30-minute subset contained 450 data points for the commercial lidar and 1,800 data points for the prototype lidar. This clarification has been added to the revised manuscript (Lines 163–165, page 8).

2.3: It is wrong that Kelberlau and Mann (2020) recommend to not fit lidar-derived reconstructed velocity component data to turbulence models. They do it in their study, are satisfied with the approach and think it clarifies lidar-specific effects of turbulence sampling.

We removed this part.

l. 245: Provide information about the "alignment condition". What range in degrees is accepted to end with 17.1 % of the data? Is this including wind from beam 3 to beam 1, downstream of the wind turbine?

It was ± 5°. We have added this sentence in the revised version: “In this paper, we restrict the application of the variance method to situations where the wind aligns (± 5°) with a single pair of opposite beams (either pair 1-3 or pair 2-4) of the lidar profilers. (Lines 265-267, page 11).

2.6: The authors should not just claim "DNV-GL has defined acceptance criteria" but refer to the source explicitly.

We have updated the manuscript to explicitly reference the source of the acceptance criteria defined by DNV GL, rather than simply stating that "DNV-GL has defined acceptance criteria." We have made this change throughout the manuscript wherever DNV GL is mentioned.

2.7: The verbal description of the quality parameters (RMSE, MAE, R2, rel. error) should be accompanied by equations that define them unambiguously.

These equations can now be found as Equations 7–10 on pages 12–13.

3.2: The description of the amount of variance included in different frequency ranges might be correct. But the conclusion that by a higher sampling rate could capture an additional percentage of the energy associated with smaller eddies is wrong. Sampling with too low frequency leads to aliasing and in a spectral display the energy from higher frequencies is folded into the lower frequency range. Instead, more focus should be put onto the relative influence of the temporal averaging caused by lower the accumulation time of the prototype lidar. Averaging does decrease the LOS variance.

As recommended, we have revised the manuscript to place greater emphasis on the influence of temporal averaging resulting from pulse accumulation time. Specifically, we now explain that the higher variance observed with the prototype lidar is primarily attributed to its shorter accumulation time, which reduces temporal averaging and better preserves along-wind variance. Additionally, we clarify that the shorter accumulation time also limits the advection-driven increase in effective probe length, thereby reducing the impact of spatial averaging, particularly at higher wind speeds. These revisions have been consistently implemented throughout the manuscript, including in the abstract, data and methods, results, discussion, and conclusion sections.

3.4.1: It is unclear why the CNR value as determined by the WindCube is not used as an indicator for the instrument noise. The median variance from spectral method for the prototype (0.0129) is also approx. 1.5 times higher than the corresponding value from the ACF method (0.0081). It is not twice as high as written in the manuscript.

Mean CNR profiles have been added to Fig. 5b (page 14) and discussed in Section 3.4.1 (lines 359–364, page 17). We also thank the reviewer for identifying an error in the median variance reported for the spectral method. As correctly noted, it is approximately 1.5 times higher than the corresponding value from the ACF method. This correction has been made in the revised manuscript (line 366, page 17).

Fig. 7: The caption should be revised to explain the different purpose of subfigures (a) and (b). Also, describe which LOS direction has been used (5, vertical?)

Fig. 7 is now Fig. 8. Here is the new title: “(a) LOS velocity spectrum measured by beam 5 of the prototype lidar (solid black), fitted using Eq. 2 with three different weighting schemes: unweighted (dashed green), low-frequency weighted (dashed red), and high-frequency weighted (dashed blue). This panel corresponds to the study focused on selecting the optimal weighting scheme. (b) The optimal scheme (high-frequency weighted) is applied to LOS velocity spectrum measured by beam 5 of the commercial lidar (blue) and the prototype lidar (orange).”, page 16.

3.5: If the mean standard deviation is 2.9% higher, the corresponding variance must be 5.9% higher. It is unclear why the authors report 7.2%?

Sorry, we have made a mistake. The mean variance of 7.2% is correct but the mean standard deviation is wrong. The true mean standard deviation is 3.5% (Line 407, page 20).

4: The discussion refers to the impact of the prototype configuration on TI but it does not critically reflect on it. What happens to TI estimates if for example the v component of the turbulence wind field becomes significant, when the inflow is not aligned with one of the beams?

We have decided to remove the discussion regarding the impact of the prototype configuration on TI, as the original computation focused on specific wind directions. This approach is too restrictive and does not adequately represent the variability encountered in practical wind power applications. As such, it does not provide relevant or generalizable insights for the broader wind energy industry. We have therefore omitted this paragraph to maintain the focus on more representative and applicable results.

I suggest reducing the discussion of the potential of the prototype lidar for floating lidar systems to one sentence since floating lidar systems are not within the scope of this study.

We agree with the reviewer that floating lidar systems are outside the scope of this study. Since it proved difficult to condense the discussion of the prototype's potential for floating lidar systems into a single sentence that integrates well with the surrounding paragraphs, we have decided to remove the paragraph entirely to maintain clarity and focus.

Technical corrections:

l. 92: "True North" is wrong here because the lidar is rotated.

You are right. We have modified this.

2.2.1: 450+1800=2256? What happened to the remaining 6 intervals?

This is 2256 30-min subsets. For the commercial and lidar each subset contains 450 measurement points. For the prototype lidar, each subset contains 1,800 measurements. We have modified this part to make it clearer for the reader. (Lines 163–165, page 8).

2.6 and other occurrences: DNV-GL does not exist anymore. They are named DNV now.

Each occurrence has been corrected.

2.6: Refer to Table 1 and include availability thresholds.

This table is now Table 2 which includes information on availability threshold. We now refer to this Table in the text (line 278 and line 281, page 12).

l. 298: Replace "almost similar" by "similar"

Done

l. 358: Replace "bin-averaged" by "wind speed-binned"

Done

l. 408: Replace "relatively slight" by "slight"

Done

Reviewer 2

General comments

Thanks to the authors for taking the time to profoundly revise the manuscript. Narrowing down the scope and including sonic data really improved the quality of the discussion. There are minor revisions that are advised before publication.

We sincerely thank the reviewer for their thoughtful and constructive feedback throughout the review process. We appreciate your recognition of the improvements made, particularly regarding the inclusion of sonic data and the refined scope. Your comments have been very valuable in enhancing the clarity and quality of the manuscript.

Specific comments

- L37: is there a way to better define the intra-beam effect to include also the time-average correctly described next? Something like "probe-time averaging".

Thank you for your suggestion, "probe-time averaging" is indeed appropriate. We have incorporated this term into the revised version: "The intra-beam effect refers to a probe-time averaging phenomenon occurring within the lidar probe, leading to an underestimation of turbulence metrics." (Line 45, page 3).

- Eq. 1 seems different from Eq. 19 in Kristensen et al. 2011. Please add additional references or a brief derivation.

You are correct, Eq. 1 is indeed different from Eq. 19 in Kristensen et al. (2011). We now include the full mathematical derivation of Eq. 1 in the supplementary material to clarify this distinction.

- L214: "However, this method performs correctly only if the range in which the turbulent cascade occurs is fully captured. "Is this because of the 2/3 power law extrapolation? Lenschow shows also simpler extrapolation methods that do not require any assumption on the shape of the AFC. Please clarify.

Thank you for the comment. We chose to focus on the power-law fit approach, as it aligns with our analysis framework. However, to address concerns about the validity of this method, we now include a stationarity test in the revised manuscript to support the applicability of the ACF-based approach.

- L221: this is the first mention of the assumption of instantaneous homogeneity. It could be better to introduce this concept earlier, possibly in the introduction, because it is fundamental to understand inter-beam contamination.

Thank you. In the revised the version, we now mention the instantaneous homogeneity in the introduction: "This effect is particularly relevant in the context of the assumption of instantaneous homogeneity, which underlies multi-beam lidar measurement techniques" (Lines 39-40, page 2).

- L245: what is the tolerance around the nominal wind direction to consider it "aligned"?

The tolerance was ± 5°. We have added this sentence in the revised version: "In this paper, we restrict the application of the variance method to situations where the wind aligns (± 5°) with a single pair of opposite beams (either pair 1-3 or pair 2-4) of the lidar profilers (Lines 265-267, page 11). Also, we have added a wind rose to illustrate the wind sectors selected for the present analysis, indicated by the gray shaded areas in Fig. 4 of the revised version.

- Fig. 4a: was there any consideration on the statistical or sampling error when evaluating mean wind speed profile? If statistical error bars were added to the mean profile (e.g., through bootstrapping, possibly circular) we may find out that the profiles are statistically indistinguishable. I doubt DNV does not require any statistical significance test.

Thank you for your comment. We have updated the figure (now Fig. 5a) to include error bars representing the 95% confidence interval, calculated using bootstrapping. This addition illustrates the statistical uncertainty of the mean wind speed profile and allows for a more robust interpretation of potential differences between profiles.

- Section 3.4.1: The application of AFC requires stationary data. If this requirement was enforced, please explain how. Otherwise, clarify that the larger scattering in the AFC method could be due to the presence of non-stationarity in the data.

Thank you for your comment. We have added a stationarity test using the Augmented Dickey-Fuller (ADF) method to assess the stationarity of each 30-minute subset and, consequently, the validity of the ACF-based approach. This test is now described in Section 2.4.2 (Lines 234–237, Page 10), and the results are presented in Section 3.4.2 (Lines 373–379, Pages 17–18).