The updated version of the manuscript now focuses exclusively on the increased sampling rate, for which a 47-day dataset was collected. This dataset is accompanied by reference turbulence measurements provided by a sonic anemometer installed on a nearby met mast. The lack of reference data in the original manuscript was another concern raised by you and the second reviewers. In this revised version, we incorporate the sonic anemometer dataset for more robust comparisons.

The paper now addresses the impact of the increased sampling rate on data availability and key performance indicators (KPIs), such as the slope of the scatter plot between the mean wind speed measured by the lidar and that measured by the reference, the correlation coefficient (R^2) of the linear regression of the scatter plot, and the mean absolute difference in mean wind speed.

We then narrowed the focus to examine the effect of the increased sampling rate on the variance and standard deviation measured by both lidars, as these values are used to compute turbulence intensity (TI), the most used metric in the wind power industry to quantify turbulence. In the first version of the manuscript, comparing multiple turbulence metrics (such as dissipation rate and integral length scale) created confusion, so we chose to concentrate on variance and standard deviation. These estimates were corrected for the variance of instrumental noise, which was quantified using two methods: a spectral approach and an autocorrelation approach, as suggested by the second reviewer.

We estimate that the updated version of the manuscript is approximately 90% revised compared to the original version.

General comments:

In the manuscript, two modified prototype versions of the WindCube vertical profiling pulse wind lidar are compared to the default WindCube v2.1. The authors look at one version featuring four times faster sampling rates with accordingly shorter accumulation times and a second version with reduced pulse length resulting in shorter probe volumes. They assess the impact of the modifications on data availability, mean wind speed, noise level, standard deviations of turbulent velocity fluctuations, integral length scale of turbulence, velocity spectra and dissipation rate.

The study is of high importance for the field because profiling wind lidar suffers from their limited ability to measure turbulence accurately. Therefore, every effort to increase this ability is sought after. The experimental setup, modifications to the lidar units, and methods for assessement of the results are overall well described.

In the overall acceptable introduction and the good and well-written description of the methods, Thiebaut et al. should be more accurate in some of their claims (see specific comments).

We have revised the introduction, which, as you pointed out, was previously acceptable but could benefit from more clarity and precision.

- The experiment with the WindCube with reduced probe length is only 4 days long. This very short trial length minimizes the value of the conclusions drawn from it and is a major drawback for the study.

We acknowledge that the short duration of the experiment with the WindCube and the reduced probe length (only 4 days) is a significant limitation of the manuscript, as pointed out by you and the second reviewer as well. In response, we have removed the analysis of the reduced probe length in the updated version of the manuscript. This is an ongoing work that is not ready to be published.

- The manuscript would benefit from a more thorough explanation of the theory behind the tested modifications. For example, it remains unclear how the increased sampling rate of the modified lidar leads to a reduction of noise, although every LoS measurement will have a higher potential for noise if accumulation times are reduced.

We acknowledge this mistake and agree that our initial comparison of the power spectral density of noise was not appropriate. The correct approach is to compare the variance of noise, which is obtained by multiplying the power spectral density by the Nyquist frequency. After correcting this, we found that the variance of noise associated with the prototype configuration is 68% higher than that of the commercial configuration. The increased sampling rate leads to higher instrumental noise compared to the commercial configuration, as the variance of noise is inversely proportional to the number of transmitted pulses, as noted by Pearson et al. (2008). Now, our results align with this expectation. We have revised the manuscript to address this point.

- The conclusions and discussion section would benefit from some more reflection on theoretical considerations and practical implications.

The conclusions and discussion but also the introduction and section 2.1 (presenting the prototype configuration) have been revised to include additional reflections on theoretical considerations and practical implications, providing a more comprehensive discussion of the findings.

Specific comments:

- l. 26: Considering a beam diameter of 3cm, only a probe length of around 17km fills up a probe volume 12 cubic meters. Thus, the statement of lidars providing averages of "up to several dozen cubic meters" appears to be a vast exaggeration. Such a statement should be backed up by references.

You are right. We have reformulated the sentence: "Anemometers estimate wind speed over a small volume of just a few cubic centimeters, whereas pulsed wind lidar profilers provide an average over a cylindrical probe several dozen meters long with a cross-sectional diameter of less than 1 cm (Fig. 1)". Lines 31-33, page 2.

No formal reference is provided for the dimensions, as they were supplied by the manufacturer during the collaboration for this work.

- l. 38: The ZX 300, one of the two market-leading wind profilers, requires one second to complete a full scanning circle. The statement that "Lidar profilers require several seconds to complete a full scanning circle" is therefore wrong and must be corrected.

We have added the word 'Pulsed' at the beginning of the sentence to exclude the ZX lidar, which uses continuous wave technology. Line 58, page 3.

- l. 80: "the WindCube lidar collects data at each location in 1 second" is wrong because it samples 5 beams within 4 seconds. So, on average the data accumulation time per beam direction cannot be higher than 0.8 seconds. And in practice some dead time for swinging the beam must be accounted for.

We are right. We don't mention anymore the sampling rate of wind speed since we are now only addressing statistics performed on LOS velocities. Here is the new version of this statement: "In its standard commercial configuration, the WindCube lidar collects data at each position for approximately 0.8 seconds before transitioning to the next. Including the time required to shift between positions, a full DBS scan is completed in 4 seconds. This corresponds to a LOS velocity sampling rate of 0.25 Hz." Lines 95-97, page 4.

- l. 89: The authors should not report an "improvement" in the method section. Instead they can write e.g., about the "modification".

We changed “improvement” by “modification”. Line 83, page 4.

- l. 89: Accumulation time is reduced by 70% to which value (0.3*0.8s=0.24s)? In line 88 1Hz and 4Hz, although there are five beams to be sampled (1Hz*5=>0.2s per beam). So, either the 4Hz or the 70% reduction are wrong.

The 4 Hz value was incorrect, but it is no longer mentioned since we now compute the standard deviation and variance only on LOS velocities, not on wind speed. Also, we removed the “70%” value since it requires confidential information to explain this specific percentage.

- l. 93: Please describe why you chose 30-min intervals, instead of 10-minutes commonly used in wind industry.

It is a common remark. A 30-minute temporal window was chosen for subsequent analysis, deviating from the standard 10-minute window typically used in the wind energy industry. This decision was guided by the need to address random errors in turbulence measurements, as highlighted by Lenschow et al. (1994). For a given averaging time T, there is a systematic difference between the true flux and the ensemble average of the time means of the same quantities, known as systematic error. This error decreases as T increases. Additionally, the error variance - representing the random scatter of individual realizations - also diminishes with longer T. Therefore, the use of a 30-minute averaging period, common in atmospheric science, is justified over the 10-minute window.

We have added this text: “The choice of a 30-minute window deviating from the standard 10-minute window typically used in the wind energy industry was informed by considerations of reduction of random errors in turbulence measurements, as discussed by (Lenschow et al., 1994). Lines 144-146, page 7.

- l 111: The short trial duration of 4 days is a major drawback of the study. Please explain why it was not possible to perform a longer experiment and how the short trial durations impeded the study.

We removed this analysis. It was just a preliminary investigation.

- 2.1.1 & 2.1.2: Please elaborate on the effect of both modifications from a theoretical standpoint. Why did you choose these modifications. What is the trade-off between duty cycle, accumulation time and sampling rate? What are the technological limitations? Why is the default configuration of the WindCube different?

Now a full section is dedicated to this (section 2.1, page 4-5).

- l. 129: Mean velocity does not have a standard deviation, but the fluctuations superimposed on the mean velocity and there is no mean velocity across the wind propagation. Please reformulate.

It is a mistake. It should have been “the standard deviation (to) the mean” and not “(of) the mean”. Anyway, we removed this sentence.

- l. 242: Please name the first and last measurement height (40m and 200m) to help the reader. Please consider to show how the 0.5% reduction is distributed along the vertical profile.

The sentence is not modified: “Moreover, the commercial configuration exhibits data availability ranging from 99.5% at the lowest measurement height, i.e., 40 m above the ground, to 93.0% at the highest, i.e., 200 m above the ground, with an overall vertical average availability of 98.2% (Fig. 4b).”. Lines 278-280, page 11.

Moreover, we have added Fig. 4b to show the vertical variation of data availability.

- ll. 244-249: IEC (2017) does not prescribe any "Best Practice" criteria. Where do the 1% and 1.5% thresholds come from?

This is a misunderstanding. The term 'best practice' refers to KPIs defined by DNV-GL for certifying lidar profilers against met masts for mean wind speed measurements. We used these KPIs to evaluate the ability of both the commercial and prototype lidar to measure mean wind speed. Please refer to Sections 2.6 (page 10) and 3.1 (page 11), as well as Tables 1 and 2, for further details.

- Fig. 8: Consider the add a titles to the figure.

The figure already includes a title in the caption, ensuring clarity without redundancy.

- Fig. 9: The legend entries seem to be wrong. "Fit - Com. lidar" ranges all the way to the highest frequency.

Thank you. It was wrong. We corrected it (Fig. 7b).

- Fig. 9: Consider to merge the two sub figures into one, since only the fitted line is different and could be compared more easily in one plot.

We did it. See Fig. 7b.

- Fig. 9: The distribution of frequency bins is different for the com. and the pro. lidar. Consider to create logarithmically spaced frequency bins for both lidars and average the spectral energy within these bins. This would make the comparison easier.

We have implemented logarithmically spaced frequency bins for both the commercial and prototype lidars and averaged the spectral energy within these bins. This adjustment simplifies the comparison between the two configurations. See Fig. 7b.

- 3.4: The noise assessment would benefit from a figure.

We have added a table (Table 3) and a figure (Fig. 8) to enhance the noise assessment, providing a clearer and more comprehensive presentation of the data.

- ll. 342-347: 0.5% reduction in data availability are not a "slight" reduction. And 1-2% deviation in mean wind speed are significant and in the order of magnitude of the total measurement uncertainty of profiling wind lidars. The authors clarify that a deviation in wind speed beyond 1.5% prohibits certification (l. 248). Here they call it “very slight bias". I understand that the manuscript focuses on turbulence estimates, still the deviations in mean wind speed must be analyzed equally carefully.

We acknowledge the reviewer’s comment regarding the 0.5% reduction in data availability and the 1-2% deviation in mean wind speed. The manuscript does indeed focus primarily on turbulence estimates, but we recognize the importance of addressing these deviations carefully. The 0.5% lower data availability in the 47-day dataset is noted, and we emphasize that over a longer campaign (typically lasting over a year), this difference may be more significant. Regarding the deviation in mean wind speed, it is now assessed using key performance indicators (KPIs), which are introduced in Section 2.6 (page 10) and further discussed in Section 3.1 (page 11). We have clarified the analysis of mean wind speed performance throughout the manuscript to ensure its significance is appropriately addressed.

- l. 350: The WindCube with increased sampling rate has a probe length of 23m that effectively acts as a low-pass filter. Still, the authors assume that an increase in sampling rate from 0.25Hz to 1Hz per beam leads to a "greater sensitivity to smaller-scale fluctuations". With a mean wind speed of e.g. 7m/s the sampling rates correspond to eddy sizes of 28m and 7m respectively. The authors have to explain why they still think, that the higher standard deviation for all three turbulent velocity components can be caused by the beneficial aspects of an increased sampling rate.

We have observed that the along-wind variance, after noise correction, is approximately 7% higher for the prototype configuration compared to the commercial configuration. This suggests that the prototype's increased sampling rate may allow it to capture the energy of smaller eddies more effectively. However, the observed improvement falls short of the expected increase, as indicated by the sonic anemometer data, which suggests that a higher sampling rate could capture 34% more energy (see Section 3.2). Additionally, as shown in Fig. 10, this effect is anticipated to be more pronounced at higher wind speeds, such as 15 m/s.

- l. 399: As commented before, the effect of increased sampling rate on mean wind speed appears significant and also the impact on data availability seems to be stronger than "slight".

We have added this:

"The increased sampling rate resulted in a relatively slight 0.5% reduction in data availability compared to the commercial configuration over the 47-day dataset. While this difference is minimal, it may become more noticeable over longer measurement campaigns, which typically last over a year for wind site characterization. Following the measurement campaign presented in this paper, the prototype configuration was installed in December 2022 on Planier Island in the Mediterranean Sea, where it remains operational. The wind characteristics derived from the full year of 2023 are presented in Thiébaut et al., (2024), including a detailed analysis of data availability. Encouragingly, up to 160m above sea level, annual data availability exceeded the 90% threshold considered best practice. Beyond this height, availability gradually declined, reaching below 70% at 220m. While this highlights an area for further optimization, the prototype lidar has already demonstrated strong performance at critical measurement heights.". Lines 408-416, page 19.

Technical corrections:

All the technical corrections have been addressed. Some of the changes may not be immediately apparent, as the text related to certain issues has been removed.

l. 71: "the mea(n) wind speed"

l. 147: "pair of (parallel) beams"

l. 167: "each...subset( )"

l. 259: "illustrated (further / in the form of scatter plots) in Fig. 4."

l. 274: "the vertical (profile) of the mean..."

l. 276: "Notably, the (deviations) were..."

l. 283: "scheme(s)"

l. 193: "three (times) higher"

l. 331: "configuration (with) increased..."

l. 334: "not exceed(ing) 5.1m/s"

l. 381: "potential(ly) improved"

Not addressed anymore in the updated version of the manuscript.

The analysis associated with this text has been removed, and consequently, the corresponding section has also been omitted from the manuscript.

l. 125: Please mention the special role of the vertical beam for this study. Many of the results shown are based on the fifth beam only.

l. 176: Noise is not(!) due to relative motion between the source and the observer, otherwise it would be the signal. Please rewrite.

Fig. 6: Please rearrange the box plots to maintain the u,v,w/b5 order.

l. 239: The absence of reduction in 100% data availability appears to be insignificant and conclusions on the effect of reduced probe length cannot be drawn from the available data. A four day period with 100% availability is not representative for a commercial measurement campaign in which the data availability of the commercial lidar is <100%. This must be reflected on in the text. What are your expectations from theory? Does the increased sampling rate reduce the duty cycle, so that there is less total measurement time?

l. 253: How does the relative deviation of 1.3% between prototype and commercial unit relate to typical measurement deviations between two random commercial lidar units? Can the results be attributed to the modifications with certainty?

ll. 266-268: Slopes below unity and positive intercepts are always expected for standard linear regression with randomly scattered data. Consider using Deming regression to assume identical random errors for the prototype and commercial lidar.

l. 295: Why are the spectral plots for reduced probe length not shown? Same for l. 305.

Fig. 11: It is unclear why the "Laminar flow" curve is considered laminar flow although it contains the highest total spectral energy of all four examples. Where would laminar flow come from in field experiments?

l. 109: It is unclear why a "50% reduction in pulse duration" leads "to a reduction in the probe length from 23m to 15m". Due to the linear relationship, the reader would expect that a 33% reduction in pulse duration leads to this reduction in probe length. Please clarify.

The probe length is not addressed anymore.

References

Lenschow, D. H., Mann, J., & Kristensen, L. (1994). How long is long enough when measuring fluxes and other turbulence statistics?. *Journal of Atmospheric and Oceanic Technology*, *11*(3), 661-673.

Pearson, G., Davies, F., and Collier, C.: An analysis of the performance of the UFAM pulsed Doppler lidar for observing the boundary layer, Journal of Atmospheric and Oceanic Technology, 26, 240–250, 2009.