Referee's comments to wes-2024-93

General comments

This paper addresses lidar improvements with a neat comparison of two new lidar prototypes with commercial systems. The finding that a reduced sampling rate is the best improvement is however poorly supported by the data.

The main drawback of the work is the lack of reference turbulent quantities to compared with. One of the systems was deployed close to a sonic anemometer but this valuable instrument is deliberately omitted. The basis on which improved turbulence estimates are claimed are mainly two and not convincing:

- Increased variance with respect to the reference lidar is by itself not indicative of improvement. As also mentioned in the introduction, lidars can overestimate variances due to cross-contamination, so how do we know that the increased sampling rate is not indeed exacerbating a positive bias in the variance? Increased variance could also come from noise, and this is has not been ruled out either.
- The reduced noise estimated from the spectra of *w* is also not compelling. Increasing sampling rate extends the spectrum to higher frequency (Fig. 9), so the behavior of the fitting can change significantly. It is also mentioned that for the commercia lidar a noise plateau was not identified, so we cannot trust noise estimates from the reference lidar so what observed in Fig. 10a can be a numerical artifact

Also, the spectral analysis shows that spectra are very noisy and therefore the results should be interpreted more carefully. For instance, the laminar flow case in Fig. 11 is very questionable as laminar flow generally does not occur in the field and also because the supposedly laminar spectrum has more variance than the turbulent spectrum.

It is suggested to profoundly revise this work to make the most out of this useful dataset:

1. Calculate the turbulent statistics form the sonic (or even cups) as well and use it as reference
2. Do not provide overall biases only, but also RMS error on a 10-minute basis or, even better, scatter plot like the one in Fig. 5 for lidar vs sonic
3. For the lidar with reduced probe volume where there is no met mast and very few data points, consider a smaller section with a lot of caution advised in the interpretation of the results
4. Evaluate lidar noise also using a non-spectral approach, like the autocorrelation method by Lenschow et al., 2000 (https://doi.org/10.1175/1520-0426(2000)017<1330:MSTFOM>2.0.CO;2)
5. The introduction could mention the effect of pulses accumulation, which is different from the sampling rate. The accumulation acts as a low-pass filter in the time domain in an analogous way as the probe average does in the spatial one. The sampling rate refers more to how quickly the lidar moves through the scan cycle, regardless of how long it takes to measure a single LOS.

These are some modifications that would bring the paper to the standards of the other publications in the topic.

Specific comments

L71: “mea” instead of “mean”

L77: is the increased sampling rate achieved through a faster accumulation or a higher pulse repetition frequency? In the second case, the maximum range may be reduced, and it should be explained.

L81: sampling rates of 0.25 Hz for wind speed may be misleading. The lidar uses a moving averaging window of 5 beams, so it does deliver a new wind speed estimate every second, but these estimates are not independent. This time overlapping effect should be made clear.

L 94: please explain what the test requirements were to consider it as “passed”.

L130: the explanation of the rotation of velocity is unclear. In general, V_x and V_y are not 0, but after rotation $v = 0$ (not V_y as indicated). Aligning the x axis to North is also not the common practice in atmospheric science, where x is W-E and y is S-N, and it may be worth mentioning this as well. Please add Fig. 1 angles and axis clearly indicated for readers that are unfamiliar with this technique

Equations 1 and 2: b terms that should be the LOS velocities are not defined.

L201: it is true that the inertial subrange is limited to the right by the viscous regime where dissipation reduces TKE, but it is also limited to the left by the integral scales that supply TKE, please add this detail.

Eq 9: the | symbol to indicate the range of frequencies may be mistaken for an integration. If a fit is instead performed in this region, it would be better to remove it and explain that it is a fitting operation in the inertial subrange.

L245: is the specification of 1% relative to the error over 10 minutes or the whole dataset? Please specify.

L255: have you considered that the increased difference close to the ground may be due to the lidar with reduced probe length being able to resolve better nonlinear mean wind shear?

L272: the increase in interquartile range cannot be automatically ascribed to a better sensitivity since it could very much be noise (instrumental or statistical). The fact that larger increases in standard deviation are seen at high altitude is also suspect in this sense, since one could expect the reduced probe length to lead to more recovery of turbulence variance close to the ground where length scales are smaller. If it happens at larger range, it could be noise not sensitivity.

Fig. 5: please add the colorbar of data density.

Fig. 6: please make the box and whisker format consistent between the two subplots.

L283: “iterative” may not be the right word, “trial and error” maybe?

L330: it is confusing saying that $\beta = 5/3$ was imposed for the dissipation energy, but then $\beta < 1$ were excluded. Is this a two-step process where first we fit β to the whole spectrum, then if it passes the check it is used for the dissipation energy with a new fit in the inertial subrange and $\beta = 5/3$? Please clarify.

L359: the integral length scale is not associated with a peak in the spectrum (not premultiplied), but it is by definition its value at 0 frequency, as shown in Pope 2020, Eq. 3.114. Please remove or rephrase.