Response to Report #2

The original reviewer comments are in black font and our detailed responses use blue font.

I have read the author's comments to both reviewers and the revised manuscript and believe that this paper is acceptable for publication after the consideration of one minor comment:

In the revised manuscript on page 3, line 25-27, a new sentence has been added that states "This is important because horizontal wind velocities are usually much larger than vertical wind velocities, and a distorted measurement of the horizontal wind speed directly affects the vertical wind speed measurement". This statement is vague and could be misinterpreted.

In the context of this paragraph, this statement is probably meant to explain the advantages of the non-orthogonal CSAT3 versus previous anemometers that were orthogonal. Yet, for an orthogonal anemometer, this statement could be inaccurate because the horizontal and vertical measurements are entirely independent. It could be argued that flow distortions due to the horizontal transducers/structure might fundamentally distort the wind flow in all dimensions such that the vertical component is fundamentally changed, leading to an accurate measurement of a distorted vertical wind component. I'm not sure if this has ever been proven or disproven. Regardless, for an orthogonal anemometer, it is not clear that there is any correlation between horizontal and vertical measurement errors.

Regarding a non-orthogonal anemometer, it is important to note that all measurement errors originate in the transducer measurements. This means that while a distorted horizontal measurement will directly affect the transducer measurements, it might not affect the vertical wind speed measurement. This would occur if the measurement error only affects the horizontal components of the transducer measurements. In this case, while the transducers measurements would be affected, the vertical wind measurement (which is proportional to the sum of the three transducer measurements in a non-orthogonal anemometer) could still be unaffected.

I would ask the authors to consider revising this sentence.

Otherwise, well done.

John Frank

Thank you for this comment. It helped us to realize that this sentence can be misinterpreted, because it does not apply to orthogonal arrangements of sonic paths. Therefore, we have modified this sentence accordingly:

"This is important because horizontal wind velocities are usually much larger than vertical wind velocities, and when using sonic anemometers with non-orthogonal paths, a distorted measurement of the horizontal wind speed directly affects the vertical wind speed measurement."

Response to Report #1

Remaining criticism:
The authors have responded well to most of the criticism, and errors have been corrected, which is good. I remain truly impressed by the PTB lidar, but I still see a few potential errors in the paper.

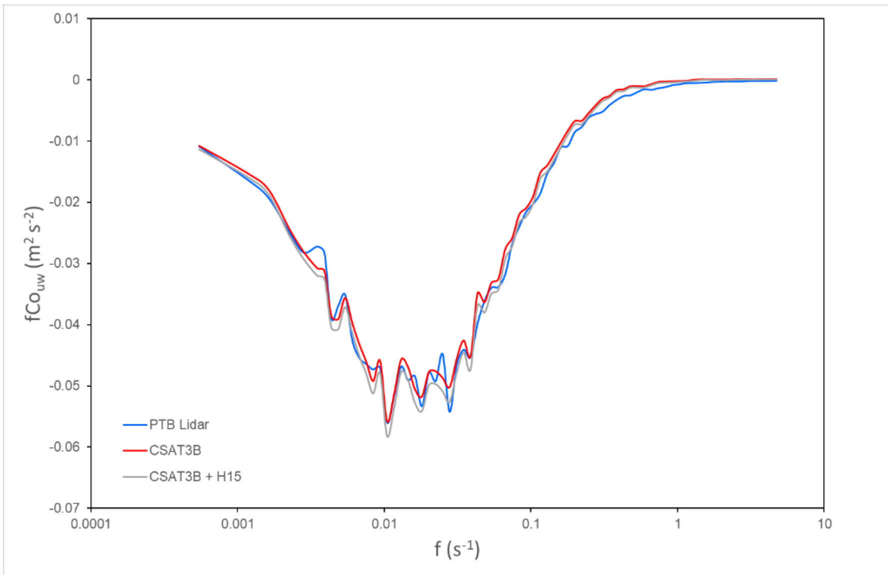
• Concerning stating uncertainties in the PTB lidar. All papers should address the main uncertainties in the presented data; this is standard scientific good practice. Given that the authors present somewhat inconsistent results (see below), it would be interesting for the readers to learn, whether there is a possible weak point. In example, the authors write in the review answer that an error as small as 0.0034deg in the azimuth is a huge problem. What is their pointing accuracy and how does it translate to wind speed errors?

We believe this is a misinterpretation of our response. A misalignment of 0.0034° is not a huge problem for the PTB lidar. It is rather something that can be clearly detected by the PTB lidar.

• Abstract:
"Analysis of the corresponding cospectra showed that the CSAT3B underestimates this quantity systematically by about 3 % on average as a result of too steep a drop-off in the inertial subrange."
The authors still claim that the difference in momentum flux between the PTB lidar and the sonic anemometer is due to the clearly seen difference in the spectra between 0.1 and 5Hz in Figure 8. However, as stated in my original review, because of the double log-axis plot, this way of illustrating error could be very misleading. In order to let the reader be able to see the frequency dependent contribution to the flux, the authors should show the premultiplied cospectrum and let the scale on the y-axis be linear. The authors have stated why they want to keep with the log-scale in Fig. 8, but this is not a strong argument. The figure is used in the context of explaining the underestimation of the momentum flux. If they want to impress with their signal quality, they should show the u-spectrum, which is a much stronger achievement, given that they measure at a near-vertical path. *In any way, they must still document that the "missing" 3% in uw flux stems from the range between 0.1 and 5Hz.* I don't believe that it does, but that rather the sum of the flux contribution in this range is much, much smaller. I believe that the difference come from the energy containing range, and if this is the case, then the H15 shows a substantial improvement of 3% in the uw comparison contrary to what the authors state now.

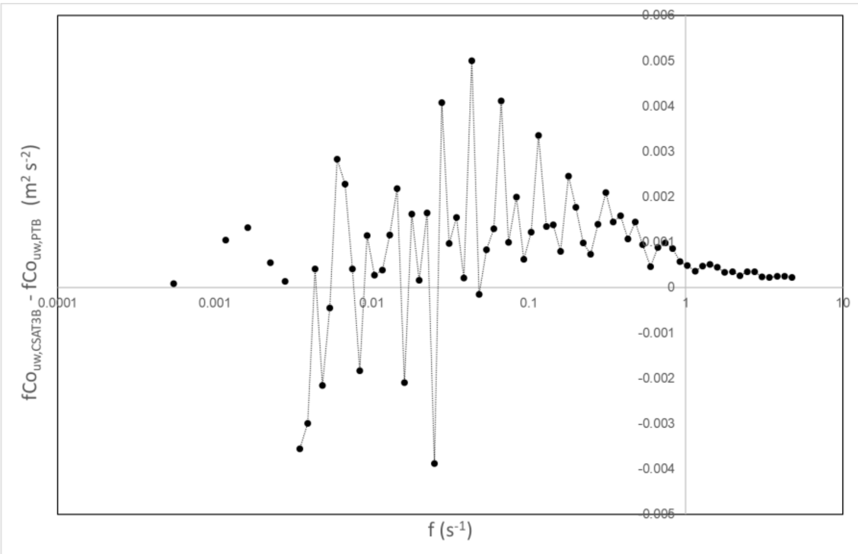
Of course, H15 cannot be used for low-pass filtering corrections, it was never intended for this use and it is utterly strange that the authors use this argument for not using the H15. Low-pass filtering effects due to path-averaging should be compensated for with path-averaging corrections, please see next comment.

It is correct that, in contrast to a double-log plot, a semi-log plot of the pre-multiplied cospectra has the property that the area under any portion of the curve is proportional to the covariance. Such a representation of the same data is presented below:



In this figure, you can also see the same behavior that we describe in the manuscript, which can be summarized in two main points:

1. The underestimation of the u'w' (and hence u*) by the CSAT3B compared to the PTB lidar stems mostly from the range of frequencies between 0.1 Hz and 5 Hz. At lower frequencies, there seem to be no systematic differences between the between the CSAT3B and the PTB lidar in the uw cospectra. For illustration of this statement, please see also the difference plot below:



2. The H15 correction leads to an increase in cospectral energy density across all frequencies, but it does not effectively correct for this cospectral loss at high frequencies, which is the main reason for the discrepancies between the CSAT3B and the PTB lidar.

Although the semi-log plot has the above-mentioned advantages, we would still like to retain the double-log version of the plot in the manuscript, because it has the advantage that the power law appears in a straight line, and this figure illustrates that the PTB lidar follow the theoretical power law much better than the CSAT3B. This is important, because it supports our choice that the PTB lidar can be viewed as reference instrument in this comparison. In order to make this point clearer and more precise we reformulated this sentence in the abstract as follows:

"Analysis of the corresponding cospectra showed that the CSAT3B underestimates this quantity systematically by about 3 % on average as a result of cospectral losses in the frequency range between 0.1 s^{-1} and 5 s^{-1}."

We have also modified the paragraph describing the cospectra in the manuscript for further clarification.

• Low-pass filtering correction/compensation for path-averaging:
The authors apply an inaccurate low-pass filtering correction erroneously, and this is not acceptable in a study where effects as small as 1% are of importance. I asked the authors to change from the Moore correction, which is an inexact approximation for path-length averaging on the vertical velocity component, to the correction published by Horst and Oncley in 2006, which is a more exact correction determined explicitly for the CSAT sonic. The answer that this is of no importance and that the authors further have applied the approximate correction for all three velocity components is not reassuring, given that Horst and Oncley (2006) showed that the three velocity components are affected differently by the path-length correction.
The authors should just do it right! Whereas the absolute size of the correction is likely negligible for large fluxes and higher wind speeds, it can be of importance for low-wind speed stable situations, and these situations were not included in the Pena et al 2019 paper. Hence, it is not obvious how large the correction is, based only on measurement height. In Pena et al (2019), the correction was applied for 6.5m and 16m but low-wind speed situations were not presented. They stated their results with and without path-averaging correction and it was clear what the effects were. This is quite different from the current study.
The only alternative to applying the path-length correction by Horst and Oncley (2006) is to apply no correction for both statistics and spectra.

I assume we all agree that any correction is just a model of the reality, and the difference between different correction models is therefore just gradual and not categorical, depending on the level of simplification and generalization. Hence, we would not agree with a statement that says that one correction is absolutely correct and the other one is inaccurate. The model of Horst and Oncley (2006) also has some simplifications and generalizations, admittedly less than the Moore correction but it is still just a model. In any case, we aim to avoid confusion and we want to present our results in a way that is as transparent as possible. Therefore, we follow the advice of this reviewer and decided to add a discussion of the results without low-pass filtering correction for path-length averaging for the turbulence statistics and also present the results of the statistical comparison in Table 1. None of the conclusions are affected because the differences due to the low-pass filtering correction are much smaller than the differences caused by the H15 correction.

• I noted that the second reviewer asked the authors to refrain from being subjective in their judgement of the results, and found this a very good point. Yet, the authors still state in the abstract that the agreement is "very good". The view of "very good" can however be challenged. In example, for wind resource assessment in the field of wind energy, systematic errors of 1-2% in the mean wind speed are viewed as problematic and the CSAT sonic shows such wind direction dependent errors if one trusts the PTB lidar; around 4% for 160 deg. and -2% for 270 deg. (Fig 6). The authors should just state their results and remove the occurrences in the abstract of "very good".
We have removed qualitative statements, such as "very good", from the abstract and now just report the results in quantitative terms in the abstract.

Further, unless the data are still affected by spikes, the result for sig_u must show the same directional dependency as that of the mean velocity. And if the sig_u shows a directional dependence, one wonders about the other two velocity components...

These larger difference in u_bar can already be seen in the scatter plot (Fig 5) and that is the reason why we investigated their wind-direction dependence. The scatter plots of sig_u, sig_v and sig_w are provided in the manuscript (or in the appendix), and they do not show such larger differences for a number of data points, which would warrant further investigation.

• lines 18-19, p. 15: Comment on citation to Hogstrom and Smedman (2004). If the authors read the whole Hogstrom and Smedman paper, it should be clear to them that the statement regarding transferability from wind tunnel to atmosphere is based on a speculation and not a proven result. Whereas wind tunnels for sure have their limitations, I believe that both Hogstrom and Smedman (2004) as well as the current authors are mistaken in their strong rejection of their usefulness. Whereas the difference in drag on cylinders in laminar and turbulent flows is indeed well-known, it should be stressed that wind tunnels are not entirely laminar, but flows in wind tunnel contain a large amount of very small scale turbulence. And it is the small-scale turbulence with similar or smaller length scale than the diameter of the cylinders that matters for the drag on the cylinder. We will for sure not settle this issue here, but since the statement is strong and of very high importance for people working with wind tunnels, I would recommend to cite carefully and correctly. However, I agree with the authors that conclusions stated in the abstract of papers should be citable, so technically, the authors have a right to cite as they do.

It is clear that the results from this field intercomparison between a sonic anemometer and a 3D Doppler lidar partially disagree with the results of from wind tunnels. The field intercomparison indicates much smaller flow distortion errors. The question is which type of experiments has more validity in principle. Since this lidar experiment was conducted under typical field conditions, where sonic anemometers are normally deployed, we are confident that these results obtained in the field are more reliable. This does not mean that wind-tunnel experiments are not useful in general, and we do not write this in the manuscript. For example, we have also conducted a wind-tunnel experiment in a previous study to assess the performance of the PTB lidar against an LDA, which is no problem at all because the lidar is free of flow distortion anyways and the comparison can therefore be assumed to be independent of the Reynolds number. We also agree that there are ways to introduce turbulence in wind tunnels as well. Nevertheless, the Reynolds numbers are still smaller by several orders of magnitude. Therefore, we believe that this is the best explanation of the observed discrepancies between wind-tunnel and field experiments. For clarification, this senctence has been added to the manuscript:

"Therefore, we believe that this explains the differences between our field study and previous wind-tunnel based and LES-based experiments, and we expect that the field experiment has more validity in principle, since sonic anemometers are normally used in the field."

What we actually also want to say is that it is generally better to avoid a flow-distortion correction through a clever design of the sensor than relying on a wind-tunnel based correction, which might not be directly transferable to the real world. For clarification, we added a sentence along these lines to the manuscript:

"Moreover, it is generally preferable to minimize flow-distortion errors to begin with through an optimized design of the instrument, e.g. by increasing the ratio between path length and transducer

diameter, than relying on the transferability of wind-tunnel based correction models to real-world conditions."

Please note that our statements in this manuscript on wind tunnel corrections are not really that strong. We use rather weak formulations such as "might not be applicable in the turbulent free atmosphere" or "it is problematic to transfer quasi-laminar wind-tunnel calibrations". Please also note that we don't write that wind tunnels are purely laminar but "quasi-laminar".