Summary: This paper is focused on the question whether or not the CSAT3 sonic by Campbell Scientific needs correction for flow distortion, i.e. correction for how the instrument structure itself disturbs the flow it measures. The answer, given towards the end of the paper is a somewhat vague no, and concerning friction velocity, the authors conclude that the CSAT3 sonic anemometer is “accurate enough for most applications”. The big news in the study is, however, the introduction of the bistatic Doppler lidar as a new flow-distortion free reference instrument, which can measure 3d turbulence with a very small measurement volume. It is also mentioned that the bistatic lidar can serve as a replacement for commercially available wind lidars, which typically measures the wind speed at several heights, typically in the range 50 m to 150 m. When new instruments are introduced, it is necessary that their limitation and shortcomings are discussed, but in this paper, only selected signals from the bistatic lidar are presented and no limitations with regards to accuracy and measurement capability/signal quality are mentioned. Further, the effect of the extensive postprocessing of the data is not satisfactorily described.

Whereas the study represents impressive works and achievements, I find that neither of the two main topics (#1 flow distortion in CSAT3, and #2presentation of the bi-static lidar for 3d turbulent measurements) is treated rigorously enough, and therefore recommend a major revision before being accepted for publication.

The presented conclusion that the CSAT3 sonic is different from the other sonics on the market and measures sig_w perfectly without any flow distortion correction is controversial, given earlier evidence. It is hard to overlook that other carefully designed studies have come to a different conclusion, and a more balanced discussion of the presented results in the light of the earlier studies would be an improvement.

Major criticisms:

1. When applying the flow distortion correction by Horst et al (2015) (H15), the authors find that the sig_w and U are both increased (Table 2 compared to Table 1). Yet, when presenting the cospectra in Fig 8, the covariance is decreased when applying H15 for all the low and energy containing frequencies. This is an impossible result; since the u and w components are both increased when applying H15, the absolute of the covariance must also increase for the low-frequency range. I recommend the authors to double check their algorithm and make sure that the red and grey lines have not been accidentally swapped. Since the u* comparison is improved by applying H15, the uw co-spectrum using H15 should also be closer to that by the bistatic lidar and a mistake in labelling/accidental swap seems likely.

   Further, the authors write:
   Apparently, an artificial correlation between u and w is introduced at high frequencies, which can be explained by the interdependence between u and w introduced through this correction algorithm.
   This argument is based on that the absolute of the cospectral density changes from around 0.00001 to around 0.0001-0.001 for high frequencies. Probably, if the absolute operator is removed, the authors would find that for the spectral range f > 1 Hz, where zero co-spectral content can be expected for the investigated setup, the sum of the co-spectral content is indeed zero and that deviations from zero are just noise. This very minor change to the spectrum is not an argument for not using H15.

2. LL 16-18, P 14: Regarding how the H15 correction changed the spectral ratio. In H15, it is shown that the implemented transducer shadowing model has a stronger effect on *w* than on *u*. This is consistent with the difference of the results presented in Table 1 and 2. Hence, since the H15 correction is frequency independent, the correction should lead to that the spectral ratio in the inertial sublayer is increased, but here it is decreased. Again, a very strange result, please double check.

3. When introducing a new instrument for measurement of 3D turbulence, measurements of all velocity components should be presented. Hence, the authors need to at least also show also comparisons of sig_v, sig_u and tilt angle. The latter should preferably be shown as a function of wind direction, since mismatches can result from imperfect alignment of the instruments to true vertical. Such misalignment will result in a cosine dependence of observed tilt angle to wind direction. These plots, if showing perfect agreement, can be put in an Appendix.

4. It is truly surprising - and point to near-fantastic signal quality - that it is possible to measure the horizontal wind speed from a near vertical path. In the direction of the wind vector, the half - angle to the emitter from the receiver in the presented setup is 0.95 deg for the receiver parallel to the wind direction and less than 0.95 deg. for the two other receivers. This means that the observed Doppler shifts are very close to zero and observations near-zero have been documented to be difficult to measure accurately using Doppler lidars (Abari et al. 2015). Further, any noise mistaken for a true Doppler change should introduce a strong signal, since it results from U = V/sin(theta/2), where V is the velocity along the direction defined by the receivers and theta is the angle between the emitter and receiver. At 100 m height, as sketched in Fig. 1, the angle between the emitter and the receivers would be 0.5729 deg. When measuring a 10 m/s wind speed at 100 m height, the velocity recorded by a receiver will be lower than 5 cm/s. For 100 m, a receiver misalignment of as little as 0.01 deg. will in the system lead to a systematic error of about 1.5% in the mean wind speed estimation. What is the limitation to measurement height/measured wind speed range, in the given setup? – please explain! And how is it possible that the horizontal velocities show such little effect of noise?

5. Data treatment: Before presentation, the data is post-processed in several steps and it is unclear what these steps do to the investigated signals. Please state/answer:
   a- how many spikes were removed in each instrument. In case there are many spikes in either of the instruments, please explain/discuss the reason for these spikes. How many removed spikes were maximum allowed in a 10 min. run?
   b- spectral treatment. The correction by Moore (1986) only corrects for path averaging on the vertical velocity component. In 2006, Horst and Oncley published exact algorithms for compensating for path-averaging for the CSAT3 geometry for all velocity components. For the spectral ratios in focus, the Horst and Oncley correction reduces the ratio (see Pena et al. Table 2) and it is therefore highly relevant for this study to implement the correct path – averaging correction. Please change/remove the Moore correction step.
   c- "After this preparation of the raw data, we discarded any 3-min statistics if more than 10 % of the high-frequency data were missing."

How many samples were removed using this step? I assume that most of the removed samples stem from the lidar?
d- It is unclear how the model spectra were used (fits to Højstrup or Moore)? Are variances and co-variances calculated from fits to model spectra?
e- Oertel et al. showed poorer agreement with reference observations for low wind speeds, but here very low wind speeds are included. What is the explanation for this improvement according to the authors?
f- The two instruments are compared in the instrument specific coordinate systems resulting from rotations V = W = 0; hence the lidar and sonic coordinate systems could be slightly different. This is from a technical perspective reasonable; it is hard to measure the exact position, yaw and tilt of each instrument. The lidar should however be easy to level, such that its w observations should indeed represent true vertical. Was leveling attempted? How different do the authors estimate the two coordinate systems to be? The latter question could be answered by plotting tilt angle versus wind direction for the two systems.

6. LL 24-27: Unlike previous sonic anemometers with orthogonal sonic paths, where the horizontal velocity components are measured from a pair of axes located in the 25 horizontal plane and the vertical velocity is measured by a single vertical pair of transducers, the flow-distortion effects in the CSAT3B are minimized by positioning all six transducers and their supporting structures out of the horizontal plane
This is an oversimplified statement. All sonic anemometer designs suffer from flow distortion and the extent is highly dependent on the wind direction and attack angle relative to horizontal. The arrangement of transducers in the CSAT3B sonic is to my knowledge designed for a low flow distortion effect on the *horizontal* wind components and a minimized effect of white noise on the vertical velocity component. In a flux measurement, the greatest contribution of the measured flux comes from incidents of high angle of attack on the instrument (Gash and Dolman, 2003), and for high angle-of-attacks, the effects of transducer shadowing increases in the CSAT3.

7. Fig. 6: This is an interesting result, possibly indicating that the flow accelerates through the measurements volume of the CSAT3, leading to an overestimation of the horizontal wind speed. H15 hypothesized that transducer shadowing was the major cause of flow distortion in the CSAT3, which can only lead to an underestimation of the velocity, so here we are potentially looking at another major cause of flow distortion. But the presented result also leads to more questions: what kind of fluid dynamics process describes a wind acceleration without affecting all the velocity components? In other words, can an instrument that measures the vertical component perfectly (as it is claimed in this study) measure the horizontal velocity imperfectly? There could be several reasons behind the mismatch in the results; small inaccuracies in the lidar measurement height or a larger focus volume than anticipated, effects from post-processing of the data, or other inaccuracies in the optics of the lidar.
8. P. 15, LL 27-30, "In our case, the H15 correction even results in improved u_* values, but the ensemble cospectrum shows that this improvement occurred for the wrong reasons. In consequence, the observed behaviour of this correction for u_* may very well be site-specific and not universally transferable. Moreover, as stated by Wyngaard (1981), such corrections are problematic because they violate conservation of vorticity and can therefore not generally be recommended.". As stated above, I doubt

that the very small change in the inertial subrange will lead to a significant contribution of the u*. Moreover, the citation to Wyngaard is very strange. The mentioned tilt correction in Wyngaard (1981) has to do with the double rotation (which is used in this study) to V = W = 0. In my understanding, Wyngaard in 1981 stated that flow distortion effects cannot be avoided regardless of coordinate system. Please point to the exact place in the paper, where Wyngaard stated that flow distortion correction cannot be safely applied or remove the citation.

9. Concerning isotropy and spectral treatment. The authors should apply stricter criteria to ensure isotropy. It is not enough to select an interval where the premultiplied spectra are flat. In Pena et al (2019), a stricter selection is suggested (for example, co-spectral density should be close to zero). Moreover, it is more correct to ensemble average the spectra based on wave number instead of frequency, since the spectral content changes as a function of wind speed. The cited Stipersky and Calaf (2018) does *not* dispute that the spectra in the inertial subrange shows isotropy, but they show that this cannot be assumed for all spectra. Hence, their result is well aligned with the methods chosen in Pena et al. (2019), but not well aligned with the method for calculating spectral ratios in this paper.

Minor comments:

- L. 14, P. 6: Regarding synchronization: It is of no importance that both instruments were sampled in UTC, what matters in a 1:1 comparison is that the signals are simultaneous. Were the two instruments logged on the same data acquisition system or were both systems synchronized to GPS time? Or was synchronization attempted via the measured time series?
- Since the path-averaging is compensated for (at least for the vertical component), why do the authors expect low-pass filtering? And how should a time constant be interpreted in relation to the source of the low-pass filtering effect (path length!)?
- Boom length and mast diameter?
- Please merge Table 1 and 2, for an easier overview of the results.
- I am surprised that the CSAT spectra show no sign of noise in the high frequencies. Is it because we are looking at the ensemble *model* spectra rather than the observed spectra? Please show the original spectra.
- LL 18-19: Perhaps the numerical simulations were not turbulent enough, so that wake effects are exaggerated, as it has also been found for wind-tunnel calibrations (Högström and Smedman, 2004).
  It is *not* shown in Hogstrom and Smedman (2004) that wake effects are exaggerated in the wind tunnel. This is a speculation on the side of the authors. The early Gill sonics were not as good as the sonics of today. For example, three consecutive runs in a wind tunnel showed considerable scatter for the same sonic anemometer (Fig. 3 in Mortensen and Hojstrup, 1994), possibly due to changes in temperature. This could also have been a reason for why the Hogstrom and Smedman (2004) observations differed in wind tunnel and atmosphere. In any case, please correct the citation to be more precise.
- Last sentence in Abstract: We also found that an angle-of-attack dependent 25 transducer-shadowing correction does not improve this agreement effectivel because it leads to an artificial correlation between the *three* wind components and therefore severely distorts the shape of the cospectra.
  Only two velocity components are shown.

- LL 32, p 4: Please provide a reference for proof of the statement that the lidar measures the back scatter from each single aerosol (and not all aerosols in the measurement volume).

REFERENCES:

Abari, C. F., Pedersen, A. T., Dellwik, E., and Mann, J.: Performance evaluation of an all-fiber image-reject homodyne coherent Doppler wind lidar, Atmos. Meas. Tech., 8, 4145-4153, https://doi.org/10.5194/amt-8-4145-2015, 2015.

Gash, J.H.C. & Dolman, Han (A.J. (2003). Sonic anemometer (co)sine response and flux measurement I. The potential for (co)sine error to affect sonic anemometer-based flux measurements. Agricultural and Forest Meteorology. 119. 195–207. 10.1016/S0168-1923(03)00137-0.

Horst, T. W. and Oncley, S. P. (2006): Corrections to inertial-range power spectra measured by CSAT3 and Solent sonic anemometers, 1. Path-averaging errors, Bound.-Lay. Meteorol., 119, 375–395.

Mortensen, N. G., & Hojstrup, J. (1995). The Solent sonic - response and associated errors. In Ninth symposium on meteorological observations and instrumentation. Preprints (pp. 501-506). Boston, MA: American Meteorological Society.