General Comments:

The article is a strong contribution on a novel topic that deserves publication in AMT. The authors first verify that a dual-doppler lidar situated in a wind tunnel can recover adequately the mean and spectral character of hotwire data, given certain limitations due to the pointing direction and volume averaging of the lidar. The discovery of a 0.5 coherence level as the frequency threshold of valid lidar spectra rather than a threshold based on Taylor's hypothesis appears novel. They then describe the validation of a new model for the lidar-derived spectrum. The spectral model with misalignment and noise is clearly an improvement over the plain Lorentzian model without misalignment (c.f., Figures 23-25), and offers a new path to get an unfiltered u spectrum from a Doppler lidar measurement. Despite the presence of an admittedly large calibration error that casts doubt on some results, the overall objective of the article is still met.

I do have three main concerns in addition to the line by line comments, which elaborate further:

1. The repeated reference to better correlation for TI=22% than TI=3%...I believe the RMSE is lower for the TI=3%, and the better R^2 correlation of the high TI case is just a function of the range of velocities used in the linear regression. The authors' explanation about relatively higher energy content in the low frequency range (that the lidar can resolve well) for the higher TI case is understood, but the reviewer wonders if this is really a critical piece of the puzzle or not. See my comments further below.
2. The discussion and reasoning around equation 11 is a little bit sparse and not completely intuitive.
3. Dual doppler vs single doppler. It is not entirely clear to me that the dual doppler data belongs in this paper, since the main results in this article only use WS2 (because the spectral transfer function is for a single doppler only). I would ask that the focus on the dual doppler reconstruction be removed or justified better. Moreover, the authors state in more than one place that the dual-doppler reconstruction is better than either of the individual wind scanners for the u component of velocity. I don't think this is supported by Figure 17/18 or Table 4. If the dual doppler is NOT better, this doesn't overshadow the usefulness of a dual-doppler technique which can resolve two components of velocity, but the text appears misleading about the u component results. (I also noticed that the dual-doppler results are not even mentioned in the abstract, which further makes me wonder: why even have the dual doppler results in this paper?)

Specific Comments:

Abstract:

- No comments.

Introduction:

- Well written.

Methodology, Part I

- Well written.

Methodology, Part II

- Lines 173-174: does the work of Sjoholm et al. and Angelou et al. specifically describe the higher frequencies in the lidar spectra as white noise? While it may seem obvious to some, there is not strong justification given for why white noise (i.e., the Gaussian distribution) is chosen. As this is a main advance of the paper, I think a little more information is warranted.
- Line 212: it is not self-evident that sigma_n should be a function of energy dissipation rate or wind speed. Is the energy dissipation and wind speed related to the decorrelation time? Is this measurement shot-noise limited? Could you explain this more?
- Equation 11: Why is gravity a relevant variable in the dimensional analysis?

Results and Discussion

Comparisons between WindScanner and hotwire…

- Line 255: I don't know if I would put so much weight on R^2 values here. I wonder how the RMSE compares between figures 9 and 10. The larger spread of u from a turbulent field seems to be giving higher R^2 even though there is clearly more absolute variation between u_p and u over most of the range for the more turbulent case. If you were to run the lower TI case at a freestream velocity of both 5 m/s and 15 m/s (i.e., over the same range as shown for the high TI case), the R^2 of the combined data for the low TI case would be larger than for the high TI case, right?
- Line 284-285: I understand that you are saying the small scales play a more dominant role for Figure 11 than Figure 12, which seems true based on the low frequency amplitudes of S(f). The energy content at f_cc is still more than 10 times lower than at lower frequencies for Fig. 11, though. Why don't you integrate Figures 11 and 12 from 0 to f_cc and from f_cc to infinity. See what fraction of the turbulence is not fully resolvable by the lidar and report this rather than emphasizing the difference in R^2 values, which doesn't seem as relevant to me.
- Line 289: I was expecting this line to say, "A possible reason is the insufficiency of the full-width half maximum metric to characterize the effective probe length." Don't you agree? What about the implicit assumption that the turbulence is isotropic, could this also be a possible culprit?

Analysis of WindScanner measurements of turbulent gusts

- Line 325: this is the first time you've mentioned 1 Hz averaged time series. Could you please give a brief mention of why you perform this time averaging (I assume to get out of the small eddy range that can't be resolved by the lidar).
- Line 331: Can you comment on why the green line is not the highest for the 1 Hz data. Is it that at 452 Hz, the two lidars are both filtering small-scale turbulence and thus agree quite closely compared to the unfiltered HW, but at 1 Hz, both the hotwire and lidar are on more even playing field and can both resolve all the scales?
- Line 338: The lowest errors appear to be found for the blue line not the black line, and a quick subtraction of the columns in Table 4 suggests that the mean difference between WS1 and HW is smaller than between WS and HW. Please revise this statement or justify it. Is the mean error of the dual-doppler reconstruction related to the fact that the hot-wire only measures one component? Why is validating the dual-doppler reconstruction given so much weight in this paper?

- Line 342: you say there is a “bias between WindScanner 1 and 2, which increases linearly with the mean wind speed”. This is not obvious from the plot except moving from \mu ~2 to ~5 where the gap widens between red and blue. Please revise or justify.
- Line 350: you say, “they do not have physical significance”. Please clarify your statement about physical significance as this is clearly a physical phenomenon in the flow that is being resolved by both measurement systems.
- Table 5: I wonder if the ratios of f_c/f_cc in this table are possibly more important in the long run than the 0.5 coherence observation, since in a real application of this technique, you will not have a reference instrument to calculate coherence, right? Would it be appropriate to suggest that the effective probe volume given by the FWHM maximum could be at most an order of magnitude in error based on this data?
- Line 375: I think this is a good conclusion to draw. It looks like the amplitude of the protocol-induced gust is larger in the u rather than v direction – could this be a reason why the triple repetition is being lost in Figure 22? Reading further, I see that these differences are quantified in Table 6. If you believe my argument, I think you could comment on how the fact that sigma_v/sigma_u << 100% might be related to your conclusion in line 375.

Modeling of WindScanner characteristics in the frequency domain

- Line 392: you say, “the latter curve is not valid for large-scale structures”. Just to clarify, is this because it is only derived for the inertial subrange?
- Line 415: the potential application is very interesting and seems worthwhile. You have used the word “atmospheric” twice in the last three paragraphs. From the introduction of the paper, I was under the impression that you want to use the dual lidar technology in wind tunnel studies of wind turbine configurations? Could you clarify here (and in the abstract/introduction) if your aim is for wind tunnel or field measurements (or both)?

Conclusion

- Line 427: In reference to 1.1%, you say that the dual doppler gives “lower spread”. However, the 1.1% comes from an analysis of mean error, not scatter. Could you clarify this wording?

Technical Comments

- Figure 1 might be more useful if it included a zoom in of the nozzle with the active grid.
- Line 68: add “to” before reproduce
- Table 2: I think it would be appropriate to give the names of the variables in Table 2 and not just the symbols.
- Line 140: you mention that L is the probe length twice.
- Figure 11/12: Could you note in the caption that f_cc will be defined later in Section X…?
- Lines 327-328: no need to describe what the different colored lines mean since it’s in the figure.
- Lines 346: no need to describe the line colors in the text.
- Line 427: “down to -1.1%.” Can you say “within 1.1%” instead to be more precise?