Monday, September 19, 2022

Dear Referee Number 2:

Thank you for the well-articulated and helpful comments on our article. Please see below our responses in green to the items you mentioned to be revised.

General Comments
- In contrast to the two other techniques the Machine Learning approach not only filters spectra, but processes the whole way down to u_LOS. It is not really a QA/QC technique anymore but rather a full retrieval, hence I consider statements like "Our work compares three QA/QC techniques, including conventional thresholding, advanced filtering, and a novel application of supervised machine learning ..." (l. 10) as inaccurate. This comment is well taken. In truth, all three methods described in the Processing Techniques section include both QA/QC *and* mean frequency estimation. From that perspective, all the techniques could be considered full retrievals, and we have synced the terminology throughout the article to address this point. Specifically, references to "QA/QC techniques" have been modified to be "processing techniques" except for the several cases where the QA/QC process in particular is being discussed. The first time the term "processing techniques" is introduced, it is now described as "…processing techniques (i.e., full retrievals including both quality assurance/quality control and subsequent parameter estimation)".

Why did you not choose to just using ML for QA/QC to produce filtered spectra and then run the peak detection on these spectra as you do with the other approaches? This would make the approach more general and give more insight on what the model actually does, hence allow for more targeted improvements We agree that your suggestion makes good sense and would have been a possibly more scientifically rigorous approach to this study. Unfortunately, we set the architecture of the ML process early in our work to process all the way to the QoI. We have added a line in the Discussion section that addresses your point, "It is also noted that more targeted improvements to the machine learning technique might be possible if the technique was designed to produce intermediate filtered spectra rather than only estimating the final QoI."

- The paper could make more clear right from the abstract and the introduction that the main aim of the current work is to get rid of solid interferences while keeping the data availability high. Indeed Fig 11 and Tab. 3 show that for the other cases already the very basic thresholding approach is enough to achieve comparable results than for the two more advanced techniques We agree with your point that the primary outcome of this work is to remove solid interference while keeping data availability high. In the third sentence of the abstract, we added the phrase "especially that due to solid interference" to differentiate that there was greater success in reducing error from solid interference than in reducing error from amplitude noise, which was the other source of uncertainty targeted by our work. Further, in the next sentence we added the underlined words: "Our work compares three QA/QC techniques…based on their ability to reduce uncertainty introduced by the two observed non-ideal spectral features <u>while keeping data availability high</u>". We also highlight the data availability aspect with a new clause in the Introduction section, "we compare $\tilde{u}^{los}$ to corresponding values measured from a meteorological tower co-located with the lidar focus point <u>while also tracking data availability associated with the different QA/QC processes</u>.

We also made a change in the Conclusion section where it was added: "However, the processing techniques worked to mitigate uncertainty due to two other sources, amplitude noise and solid interference, while keeping data availability high, and most of the benefit of the higher-fidelity techniques stemmed from the reduction of error from solid interference." We believe these edits highlight the primary outcome of the work while not entirely removing reference to the ability of the processing techniques to sometimes reduce error stemming from amplitude noise, which was demonstrated to a small degree in Figure 11.

Specific Comments

- 15: please re-word "overlapped meteorological tower" to something like "on-site meteorological tower" or "meteorological tower within the sampling volume" We have changed this to "adjacent meteorological tower within the sampling volume".
- Sect 2.2: This description is quite long for that it actually principally follows Herges and Kayantuo, 2019. Please consider a more concise formulation focusing on potential differences to what is presented in the above-mentioned reference. Nevertheless, the illustrations Fig. 2 and 3 are rather valuable, I wouldn't skip these. This point is well taken, and we have reduced the length of this subsection by almost a page including eliminating the most "in the weeds" discussion and figure surrounding the final filtering step. Note that we keep Fig. 2 and 3 while removing Fig. 4. This reduction brings the total length of the subsection to be similar to that of the following subsection on the machine learning technique, which we see as appropriate since the article serves equally as a validation of the advanced filtering technique as well as the machine learning one.
- Fig. 2: I found it a bit confusing to have the most "raw" data in subfigure c) while excerpt spectra of it are in b) and processed median LOS winds in a). If I didn't miss anything fundamental, I would prefer to change c) <-> a) Please see Figure 4 in the new manuscript. We agree with you and have switched the order to what you describe and updated the text and caption accordingly.
- Fig. 3: Please leave a reference to the original spectra in Fig. 2b) in the caption of Fig 2) The authors are not quite sure what the referee is asking. These are the original spectra that came directly out of the instrument save the offset and scaling described by the axis label.
- 301: you deliberately exclude double-peaked spectra for training what is to be expected present for wake situations. Why this choice? How to explain that the Machine Learning results fit the anemometer data that well in Fig 10d)? Double peaks occur only when very steep gradients occur near the focal point. It is not expected to ever see double peaks for inflow cases, which are the focus of our study, because the atmospheric shear is usually not so severe. The referee rightly asks about double peaks in the waked situations. We would indeed expect to see double-peaked spectra near the edges of the wake (i.e., at the steepest part of the wake shear layer). This occurs because the nearfield of the probe volume senses the inside of the wake while the farfield of the probe volume senses the outside of the wake and/or freestream. This effect is most pronounced when the focal length is longer than our 2.5D example case, but it still can occur at the 2.5D position. This is a limitation of the study in its current form. Related to the good fit of the Machine Learning data in Figure 10(d), these data are all at δ=0.2°, so the lidar beam is not intersecting the wake edges for these cases. One would not expect to see double peaks here, therefore. Figures 16 (a) and (b), on the other hand, show data near the wake edges (i.e., 18

and 45 m), and there is generally a small increase in error for the Machine Learning technique compared to the Advanced Filtering one, though the error bounds are large, as noted in the manuscript.

- I would prefer to have the experiment site description (Sect. 3) before the data processsing description (Sect. 2) as I think it might ease the reading process. Especially it will also make the section about the ML easier to read. Yes, we agree with your point and have made the switch.
- 528: "This source stems from the difference..." Your explanation to this shows plausibility for this interpretation but is not sufficient to exclude all other possible sources which could cause such a bias. For this you would need to deliver a more quantitative estimate of your interpretation. Anyways, it is maybe beyond the scope of this manuscript so you might simply go for "This source PROBABLY stems from the difference..." OK, we have made this addition.
- Fig: 10: The green "x" can be misinterpreted as a fully trusted data point (instead of an outlier well handled by your method) if not carefully reading the legend. Another notation would be preferred. Yes, please see Figure 9 in the new manuscript. We wanted to keep the color green for these data points to help associate the data with the machine learning technique rather than the thresholding technique (which is colored red), so we have replaced the green "x" with a green "o" and then overlaid a red "x" on top.
- 709: Please add a reference to the out-of-distribution-detection techniques OK, please see the added reference to Yang et al., 2021.
- 734: In what you show rather than providing "higher accuracy" ML appears to provide "very similar accuracy" to the advanced filtering method. Please re-word This is a good point, and we have removed the mention of accuracy here altogether. We believe that with future tuning of the machine learning technique, accuracy COULD improve beyond that of the advanced filtering technique, but we agree that hasn't been demonstrated yet.

Technical Corrections

- 331: please include space between "3." and "Experimental Techniques" This has been corrected here and in several other places; thank you.