Reviewer 1

The authors thank the reviewer for their comments on the paper. A common theme from both reviews was the overuse of qualitative descriptions and analysis. Care has been taken to eliminate these from the text. Below you will find responses to each individual comment.

The study tries to answer the question "how well the 2D and 3D multi-Doppler measurements perform in a complex terrain?" but does not give quantitative results in detail (or not fully comment on the results). There are assumptions made without full explanation which makes the methodology unrepeatable. I believe the paper needs major changes. I would be happy to read again after the corrections.

We thank the reviewer the feedback. We have expanded the description of the methodology that we used and have made additional major changes to the manuscript which are described in detail thereafter.

- Section 2: If you want your paper to be easy to understand for everyone, even for people who does not know Perdigao experiment, I think you should make more figures showing the scanning patterns for the lidars. I think section 2 needs more then Figure 1

The authors agree that fully grasping the instrument layout can be challenging. However, it is this complexity that makes it difficult to convey the scanning strategy of each individual Doppler lidar (DL). To address the reviewer's concern, DLs that were not used in this study have been removed from Figure 1. Additionally, we added a column to Table 1 that indicates which DL's are used in each of the virtual tower (VT) retrievals. This information, together with the exact values of the azimuth and elevations of the range-height indicator (RHI) scans for each DL provide detailed information about the scanning patterns and the instrument layout. Furthermore, we have added references to the Perdigão overview paper (Fernando et al., 2019) which provides a complete overview of the Perdigão experiment and payed attention to using site names that are consistent with the descriptions in this paper.

- Page 2 Line 11: The term "large spatial heterogeneity" is not quantitative. Since the referred studies gives a range of number, I think you should rephrase.

This has been reworded to better address the issues that arise when using DL scans in complex terrain and to include quantitative statements that are guided by the cited literature. We have also made an effort to remove qualitative descriptions throughout the rest of the text.

- Page 6 Line 1: How do you justify using "linear interpolation"? Add references and comment on the contribution of the linear interpolation on the general uncertainty

The range gates of the lidars at the location of the virtual towers are comparatively dense (Figure 1). The largest distance of an interpolation point to an actual range gate is 16 m for the CLAMPS lidar, 5 m for the two DLR lidars and smaller than 10 m for the DTU lidars:

- *CLAMPS, VT1: 15.9 m*

- *CLAMPS, VT2: 16.3 m*
- *CLAMPS, VT3: 16.2 m*
- *CLAMPS, VT4: 13.7 m*
- *DLR#1: 5.2 m*
- *DLR#2: 4.8 m*
- *WS2: 8.9 m*
- *WS5: 10.0 m*
- *WS6: 9.5 m*

We have evaluated the difference of cubic spline interpolation to linear interpolation with example profiles (Figure R2) and found that differences of less than 0.1 m s^{-1} occur (Figure R3), which is well within the uncertainty we set for the lidar radial wind speed measurement. Since neither cubic splines nor linearity can be assumed for the atmospheric flow field, we decided for the simpler, computationally more efficient method of linear interpolation.

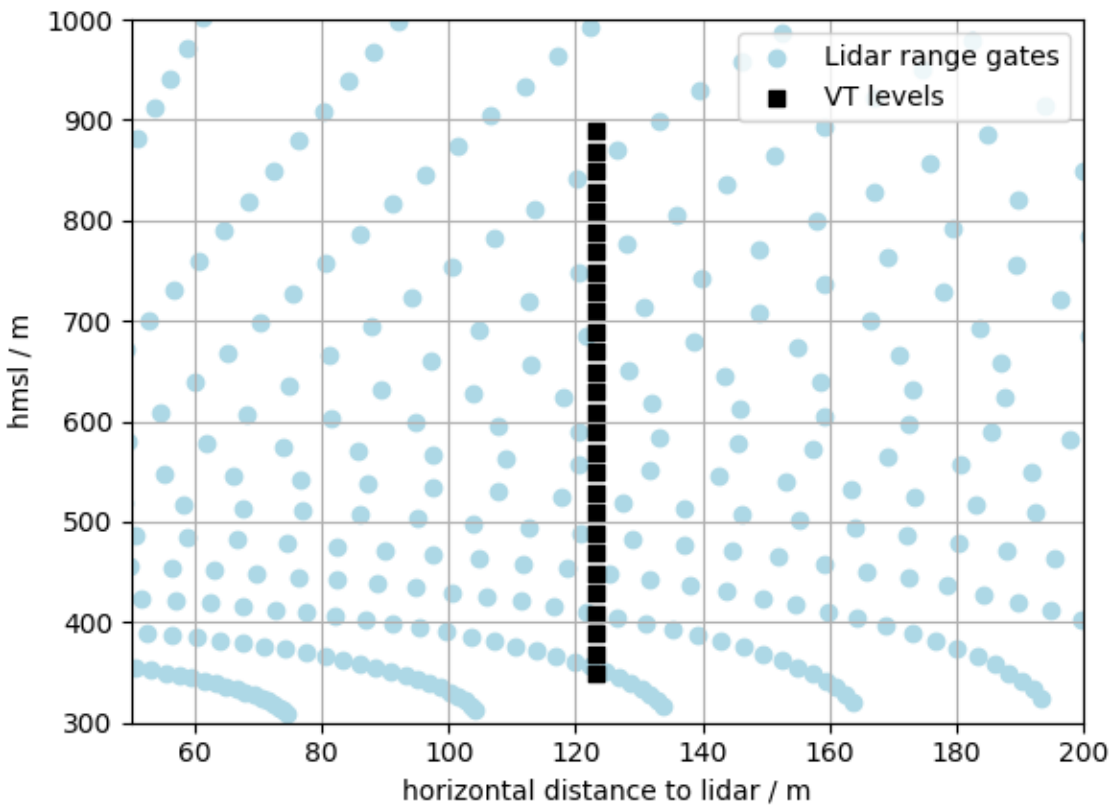

Figure R1: Visualization of lidar range gate centers of the CLAMPS RHI (blue dots) and the interpolated points for the virtual tower retrieval at location VT3 (black squares).

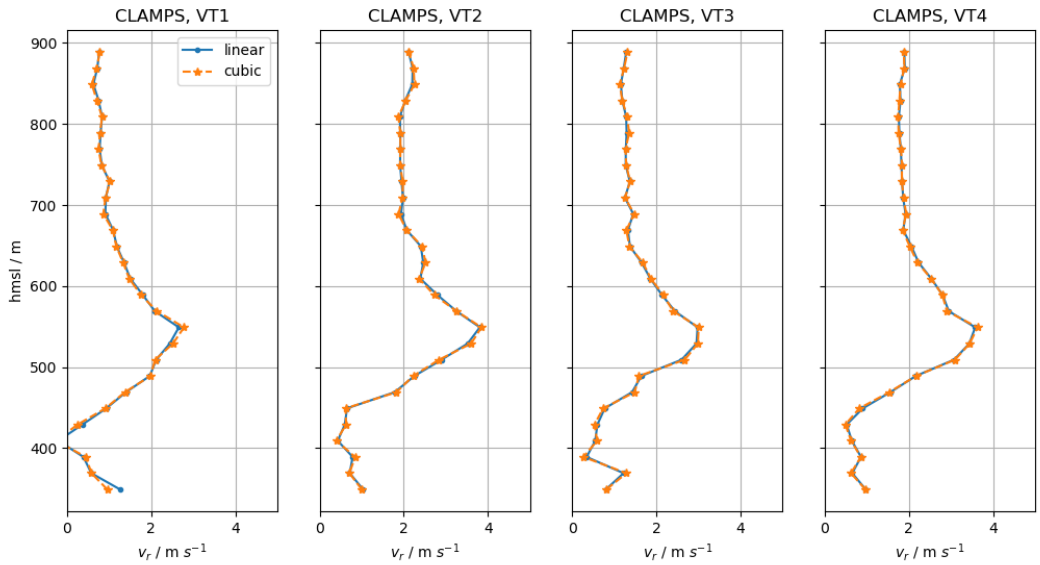

Figure R2: Comparison of interpolation of CLAMPS radial wind speed measurements to the VT locations using a linear interpolation (blue line and dots) and a cubic spline interpolation (orange line and dots)

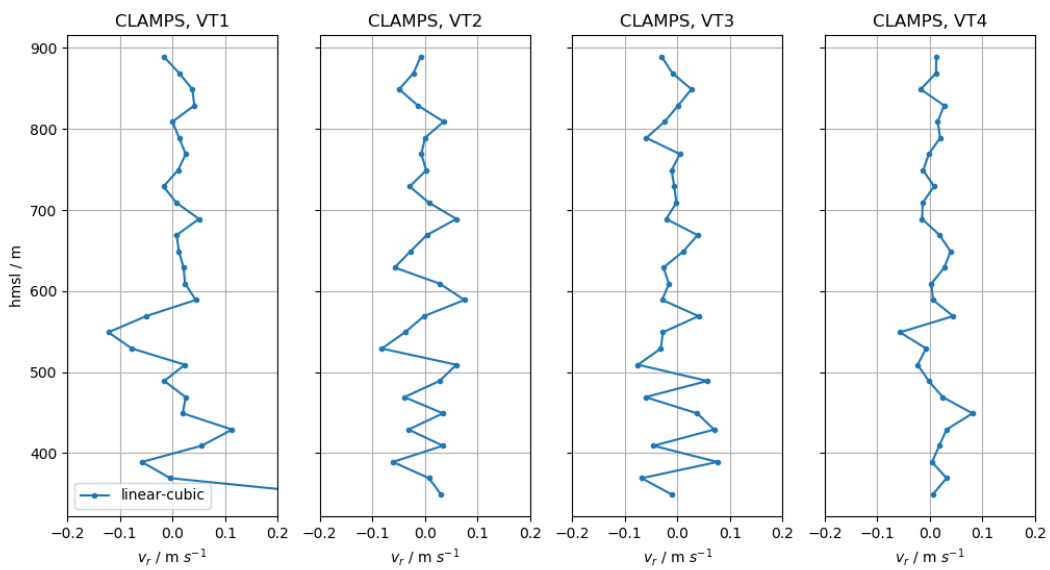

Figure R3: Difference between linear interpolation and cubic spline interpolation of the CLAMPS radial wind speeds at the four VT locations.

- Page 6 Line 20: Even expert opinions need citation. You must either write more about your assumption of 2m/s and its reasoning in detail, or you need to cite a study doing so. It is a very critical step in your method chain, and you are making the method unrepeatable by just giving an assumption for your specific case. You must clear this point.

We would like to point out that the assumed accuracy was 0.2 m/s not 2 m/s, we apologize for any confusion may have caused (the original paper stated that we used .2 m/s). This has been addressed in the text by adding citations to studies that have analyzed this problem. Though the specifications of the Windcube 200s has a specified accuracy of 0.5 m/s, Kigle (2017) found they were accurate to ~0.10-0.15 m/s in comparison to a sonic anemometer at a meteorological mast during the Perdigão campaign. Therefore, a conservative value of 0.2 m/s was used for this analysis.

- Page 9 Line 16 + Page 10 Line 5: Journal uses SI units.

Fixed

- Page 13 Figure 8: I cannot understand the difference of the vertical wind speed values above 200m. OU DL and VT3 show a big difference. Why? Can you comment?

The authors are unsure what differences are being highlighted by this comment. If the reviewer is referencing the differences in vertical velocity just below ridge height in Figure 8, this could be due to the virtual tower being considered an instantaneous measurement, while both the OU DL Stare and the physical towers are averages. Depending on how steady the flow is, there could be some differences between the instantaneous measurements and the averaged measurements (as noted in p15, lines 2-4)

- Conclusion: I think more discussion is required for the limitations of the lidars.

We mention some of the downfalls of multi-Doppler retrieval methods toward the end of the original introduction and we have slightly revised this to favor virtual towers less. Additionally, we already discussed limitations of the use of uncoordinated RHI scans in our virtual tower retrievals in the conclusions. This discussion has been further expanded to better highlight the limitations of the various profiling techniques (p18, lines 18-21)

Changes not in response to reviewer comments:

- *Figures 6, 8, and 9 in the original manuscript had an error with the error bars on the OU DL Stare profile. Instead of the standard deviation being plotted as the magnitude of the error bar, the mean was accidentally used as the magnitude. This has been fixed in the updated manuscript.*