Dear anonymous referee,

Thank you very much for your review and your recommendation for publishing this work after a minor revision. We will address all of your comments, which we have numbered 1-13, separately in the following.

1. How did the authors calibrate 3D hot-wire probe? How about accuracy of the calibration? Is there any particular reason for choosing a 3D hot-wire probe even though the lidar measurements were 2D?

   The calibration and accuracy of the hot-wire probes will be stated in the paper more thoroughly. The reason for choosing this device is simply that it was part of the standard setup in the wind tunnel. Also based on the other reviews we received, we will pay more attention to the missing vertical wind speed (w-) component. By using the information of the hot-wire we will be able to address the uncertainty it adds to the reconstructed u- and v-components of the lidar.

2. The authors worry about heating of the hot-wire probe. There have been many studies in literature where LDA and hot-wire or PIV and hot-wire were used in order to measure different flows both in and outside of wind tunnels. Why in this case the heating caused by the laser beam becomes an issue.

   Because we did not know whether or not it could have an influence, we just made sure that the laser beam could not hit the hot-wire itself, as a preventive measure. We believe it is a valid speculation that a laser beam hitting a very sensitive metal wire could interact with it and therefore we focused the beam a small distance away from the hot-wire probe.

3. The authors carried out the measurements of three different types: In the second case, the complete line was covered every 1 s with equally sampled measurements. Here, the characteristics scales of turbulence should be given for quantitative comparison. For example, what is the corresponding integral scale for this 1 s measurements. Is 1 sec statement correct?

   The turbulence scales that can effectively be measured with this setup are larger than the ones corresponding to 1 Hz, because the temporal and spatial resolution are linked. This will be mentioned in the paper.

4. Right before the results the authors give information about the wind velocity profile. Since the work compares the two system, it is of interest to see the velocity and turbulence profiles upstream of the turbines. These profiles should also be compared again the theoretical profiles like the power law instead of giving turbulence intensity at one single location.

   Unfortunately we did not execute measurements of the vertical profile during this measurement campaign. However, a vertical wind profile for both wind speed and turbulence were measured during a previous measurement campaign, under identical conditions. These plots are added to the manuscript.

5. It is not clear why they carried out the measurements at 2500 Hz for hot-wire and 390 Hz for lidar and then compared at 1 Hz. Any particular reasoning for carrying out the measurements this way?

We measured with both devices at their maximum sampling rates, to benefit the most from their turbulence measurements. However, since 2500 Hz and 390 Hz are not compatible, we first averaged the hot-wire measurements to 390 Hz to allow for a fair comparison. But because we know that the lidar probe volume averaging effect filters out small turbulence scales and does not truly resolve turbulence scales up to 390 Hz, also a comparison at a more reasonable 1 Hz was done additionally.

6. It is not clear if the figures 8 and 9 are based on instantaneous readings or the mean quantities, or turbulence. The authors call the figures correlation, but as far as seen in these figures they are individual (or mean) points from one instrument and corresponding reading from the other instrument. How accurate the time lag was introduced intothese computations? Also it is difficult to figure out the motivation for red fits in these figures, 8 and 9. Is there any particular significance? It would also be nice to see how they compute what is presented in figures 8 and 9.

   Figures 8 and 9 are regression curves of the wind speed components u and v, showing how well the lidar and the hot-wire probe correlate with each other. We believe that it is standard practice to plot this and fit a regression curve through the scattered points (see the red line). We don't understand what more motivation we should provide for doing so. The plots are based on the instantaneous lidar measurements (390 Hz) and the averaged (from 2500 to 390 Hz) hot-wire measurements. The time lag was computed numerically based on a cross-correlation function, which finds the maximum correlation as a function of time lag. The accuracy of this method is the time step in between the lidar measurements, i.e. 1/390 Hz ≈ 2.6 ms.

7. Figure 10 and 11 show very good agreement between these two signals. Can the authors elaborate effect of down sampling from higher sampling rates on the line plots?

   The lidars filter out some small scale turbulence because of the probe length averaging effect. This means that the 390 Hz regression plots are not representative of the capabilities of the lidars. When averaging to 1 Hz, all turbulent scales can be measured as well by the lidar as by the hot-wire probe and the fit becomes better. To investigate this effect further, the plots in Figures 14 and 15 were produced.

8. Discussion regarding the figure 16 needs to be further detailed. For example what is the record length for these spectral computations, and how many blocks of data are used. Even though the Taylors theorem indicates 28 Hz, lidar seems to be rolling off much earlier around at 10 Hz. Any explanation for this? In addition, the authors should write the formulation used for computing spatial averaging when finding 28 Hz. The authors should also write how they computed the spectra.

   The spectrum is based on a two minute time series of the 390 Hz data, which is split into ten blocks which are then filtered with a Hann window to smooth the spectrum. This information will be added. The fact that the slope of the lidar spectrum already deviates from the -5/3 Komolgorov rule earlier on is explained with the sentence ‚*The drop in the slope of the spectrum does not exactly coincide with the 28 Hz frequency mark, because the intrinsic Lorentzian spatial weighting function of a continuous-wave lidar extends beyond the defined bounds of the probe length, therefore also acting as a filter on lower frequencies.*‘. This could be reformulated to increase the clarity.

9. Figure 17 and 18: Here it is interesting to know about how many effective uncorrelated samples in these 1 minute recordings across the wake. Statistical accuracy changes for a fixed record length since the turbulence intensity vary a lot. What do the author mean by binned average? is this average of the open circles?

   The red and green lines indeed show the mean and the standard deviation of the measurements marked with grey circles, binned with respect to the y/D-axis. Naturally there is a high uncertainty in the measurements, because the samples might not all be correlated. Therefore the mean profile here is more relevant than the single measurements.

10. The authors note that it is hard to give any hard conclusion on the lidar's ability to measure small scale turbulent fluctuations. Previously the authors showed this when presenting the spectra. According to their approximation the cut-off frequency is about 28 Hz, which is rather low considering the probable length of the cascade, which is hard to find out since the Reynolds number is not stated as far as seen.

   We don't have hard numbers for the expected frequency range of the inertial subrange in the case of these wind tunnel measurements. However, the spectrum of the hot-wire probe measurements indicates that it extends roughly from 2 Hz to over 100 Hz, something the lidar measurements are unable to resolve. This difference between the two measurement devices is the most important point here.

   You are also indicating that we actually are able to give a quantitative measure of the lidar's ability to measure turbulent fluctuations, namely the cut-off frequency of about 10 Hz (lower than the expected 28 Hz). We will reformulate this accordingly.

11. On page 13 and in line 15, the authors writes about the small scale effect such as wake. What do they mean here? I think it is not possible to capture the small scales using this setup due to the size of measurement volume, and wake itself cannot be considered small scale.

   It is true that the wake is not a small scale itself. We meant that normally wakes introduce small scale effects such as a higher turbulence and shear on its boundary. Also we were not able to capture the really small scales in this setup. Therefore the text about this will be rewritten.

12. From the formulation given in the text, it is not obvious that the uncertainty in the y-direction has the most significant contribution. Further explanation is needed for this statement.

   The fact that only the gradient $\frac{\delta u}{\delta y}$ was taken into account when assessing the uncertainty introduced by the wake being present, is that the gradient $\frac{\delta u}{\delta x}$ is very small almost everywhere in the wind field. The reason for this is that the wake recovery takes place very gradually, but the boundary of the wake shows very steep gradients along the y-direction. This will be written more clearly in the paper for a better understanding.

13. When the author discuss about the total uncertainty, they mostly relate it to angle difference. Due to the nature of the flow, however, turbulence intensity also plays an important role in any uncertainty computation due to the statistical convergence. Toward the edges of the wake, the mean velocity drops and turbulent fluctuations as well. But the intensity can be very high? What would be the effect of this on their uncertainty calculations. Another question in relation to this one is that the wake develops downstream and velocity deficit gets smaller and smaller, and this leads to stronger demand on resolution. What would be the effect of this on the performance of lidar data, and the uncertainty. One can look at figure 21 and 22 to get an idea, but there the uncertainty is higher along the tip vortex, and wake development does not matter.

    It is true that the uncertainty analysis is mostly focusing on the lidar setup, i.e. the difference in the azimuth angle of the two lidars. The emphasis of the uncertainty analysis is the error introduced on the reconstructed wind speed components by the dual-Doppler lidar reconstruction. We do not have reliable measurements of how the turbulence intensity varies over the wind field and in the complex case of the wake, this is surely an important issue. However, this is not regarded here because it is out of the scope of our paper. Therefore a simplified model was established to express the uncertainty added by the wake's deterministic properties.

    As stated before, the uncertainty of the dual-Doppler reconstruction is the most important aspect considered here. The wake development does not have a direct effect on the **absolute** value of the uncertainty. If you express it as a percentage of the velocity deficit, the uncertainty will grow, but according to us this is an artificial effect. Therefore we chose to not express the uncertainty in terms of the wake deficit.