We thank the reviewer for the considerable time and effort put into reading and critiquing the manuscript. We believe that the revised paper has significantly improved by addressing the concerns of the reviewers. Below, we have included a response to each of the reviewer's comments, and our responses are shown in red.

1) Lenschow et al. (2000) derived the higher-order moments from the noisy lidar data in the convective boundary layer (CBL) i.e. during convective conditions. All the studies mentioned in the manuscript using the auto-covariance technique were made during convective conditions when the CBL top became quasi- stationary and turbulence was dominant. Here in this paper, the authors show results from the total observation period including stable conditions where turbulence was weak. Is the inertial subrange detected in the stable conditions from Doppler lidar measurements? Are the major part of the turbulent fluctuations resolved during these periods. The spectra shown here in Figs. 8 and 9 show that the inertial subrange was not reached in all cases.
The authors show the integral time scale in Fig. 4. Please also show the integral time scale, tint (or length scale) profiles for the respective cases with Doppler lidar measurements. It is important that the major part of the turbulent fluctuations must be resolved with the measurement system, so that the temporal resolution(dt) of the system is sufficient (tint » dt) to sample the turbulent eddies, so that the inertial subrange in the spectrum and dissipation range in the auto-correlation function becomes resolved (Wulfmeyer at al .2010; Lenschow et al. 2012; Turner et al. 2014). If all the above conditions are met, then it is possible to derive the variance profiles. It would be good to split the total data into stable, unstable and neutral conditions. For example, during stable conditions (Fig.7c), the first 4 min data shows different behavior between the lidar observation and the sonic measurement. Hence, this will also show the differences in the variance comparisons with respect to a variety of conditions and improvements.

In the updated manuscript, additional references are provided wherein the auto-covariance technique is used on routine (i.e., continuous 24-hr, averaged over 10, 15, 20, or 30 min) observations to remove noise, including during stable conditions. Since it has been used for stable conditions in several studies in the past, we determined it was important to investigate the use of it for routine observations regardless of whether or not the inertial subrange is resolved. For these types of measurements, it is not practical to manually ensure that the inertial subrange is detected. That is why within this study, we evaluate the technique under all conditions and for all 30-min periods, regardless of whether the inertial subrange is resolved. This helps determine whether or not it is useful to apply under stable conditions. We do show an example of spectra and autocorrelation (Fig. 9 a, b) where the inertial subrange is not clearly detected, and discuss it on p. 20 l. 9-17. In addition, we have made several modifications throughout the manuscript to reflect this objective, which was not clearly indicated in the previous version of the manuscript. We have also made some edits throughout the results section and added a few sentences in the conclusion to clarify that the technique may lead to underestimates when the inertial subrange is not resolved, and that it should only be used if the noise contamination is sufficiently large.

We agree that it would be best to split the data into stable, unstable, and neutral conditions for analysis, but with only a 2-day period available and the conditions predominantly neutral or slightly stable, the statistics for each condition would not be significant (in particular for unstable or strongly stable), especially for comparison to each other.

2) Here the authors choose 30 min period for the analysis of the vertical velocity variance. In general most earlier studies use 2 hour time period for the analysis of higherorder moment data in convective conditions. It would be interesting to compare if any differences would be found if the same analysis is performed using 1 hour and 2 hour data sets for the variance estimates. This will also make the study more robust.

In the initial preparation of the study and manuscript, we internally debated the averaging time to use, considering 10-min (wind energy standard), 30-min, or 1-hr. We settled on 30-min. This was used so that the statistics would be more robust and sampling errors would be smaller than 10-min. Additionally, the PBL structure often varied quickly (see Fig. 2 in manuscript) with increasing/decreasing wind speeds. Thus, with a 1-hr averaging window, these variations would be averaged out. We have provided several figures below of the main results using different averaging times (10 min vs 1 hr). We did not try using a 2 hour averaging window since that would greatly reduce the limited number of points for comparison. No matter the averaging time used, the main conclusions of the study are not affected(i.e., autocovariance technique improves estimates by decreasing variance when variance is small removing noise, and increasing it when it is large and scales are not resolved). The main difference is that the R^2 values slightly decrease for longer averaging times and the fitting generally becomes closer to a 1-1 agreement, albeit with less data points. This confirms that sampling errors are not a significant source of discrepancy between the lidar and sonic measurements, due to the short distance between sampling volumes. Otherwise, it is expected that the variances between the lidar and sonic anemometers would show larger scatter for shorter averaging times. With these differences being minor and not altering the conclusions within the scope of the paper, we feel it is best to not greatly expand the paper to discuss the effect of averaging time. However, within the updated manuscript, we have added a paragraph with a more detailed discussion on why a 30-min window was used and the effect of averaging times on the statistics presented on p. 12, l 8- p.13 l 2.

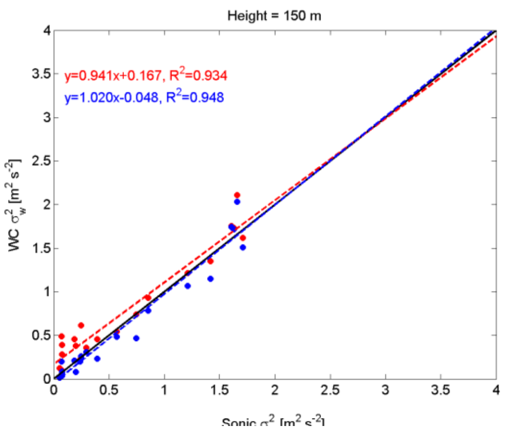

Fig. 1: Values of σ_w^2 calculated from sonic anemometer and LLNL WC data, same as Fig. 10, d in manuscript, except using 1 hr averages instead of 30-min.

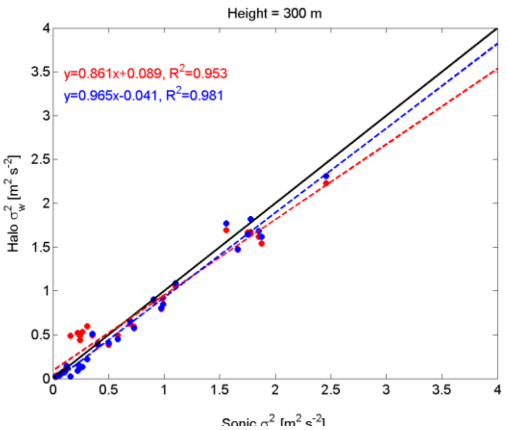

Fig. 2: Values of σ_w^2 calculated from sonic anemometer and OU DLdata, same as Fig. 10, b in manuscript, except using 1 hr averages instead of 30-min.

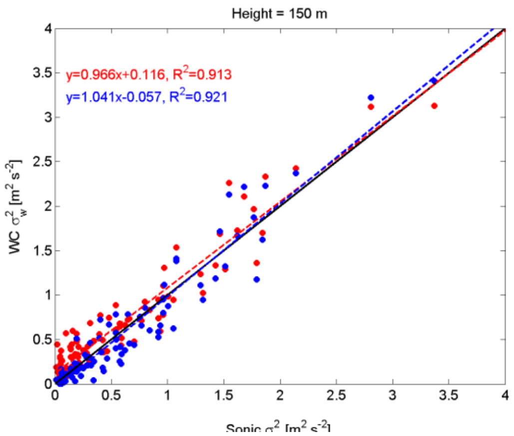

Fig. 3: Values of σ_w^2 calculated from sonic anemometer and LLNL WC data, same as Fig. 10, d in manuscript, except using 10-min averages instead of 30-min.

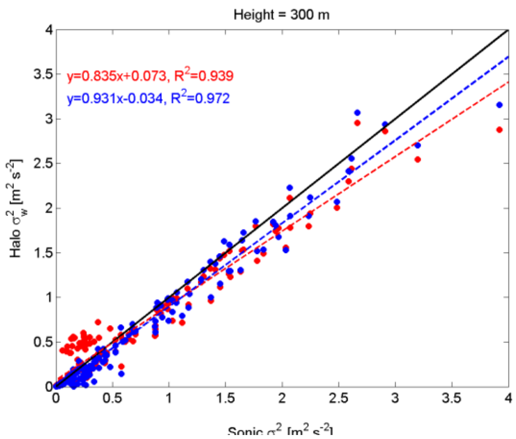

Fig. 4: Values of σ_w^2 calculated from sonic anemometer and OU DL data, same as Fig. 10, b in manuscript, except using 10-min averages instead of 30-min.

3) The number of lags needed for the extrapolation of the structure function to lag zero is given in detail in Behrendt et al (2015 and references there in). The zero crossing of the auto-covariance function appears at 2.5 times the integral scale, hence the resulting integral scale must be larger than the averaging time of the measured data. Here in this study, in Fig. 9 a shows that the zero crossing is already found at 15 lags and the tmax used is 8 s. The above condition is not met here in this case. For the other cases in the same figure larger variances are found indicating turbulent periods and zero crossing is found at higher time periods. Also, for the lower variance found in Fig. 9a no changes are found in the tmin and tmax values when compared to 9b and 9c. Hence this method of choosing the time lags is questionable? Instead the authors can verify how the differences in the lags chosen for the variance analysis show corresponding differences in the variances obtained at different heights.

Within Fig. 9 a, the zero crossing does appear at 15 sec of lag, as the reviewer states. Thus, the integral time scale (t_{int}) is approximately 6 sec (15/2.5), which is indeed larger than the averaging time of the measured data. Behrendt et al. (2015) state that the amount of lagtime used should be <2.5 times the integral scale. Thus, the lagtime used should be less than 15 sec (t_{int}*2.5). The maximum lagtime used in Fig. 9a is 8 sec, which does in fact fulfill the criteria within Behrendt et al. (2015) for this case since it is less than 2.5*t_{int}. Since the criteria is met, no changes were made to the manuscript.

The reason that the values of t_{min} and t_{max} did not change for cases shown in 9c, e is due to the fact that the LLNL WC only samples every 4 s. Thus, the amount of lag time used is discretized. If samples were taken more frequently, the amount of lag time used would indeed change between each case. This has been clarified in the manuscript.

Due to the fact that the quality of the northwest sonic anemometer data at 50 m was poor and that SNR at 200 m for the LLNL WC was generally low, we have decided not to show how differences in the lags chosen for the variance analysis vary over different heights, since that would only allow one other height to compare with. However, we believe that this would be useful to investigate in the future as certainly the number of lags will change with height, as the typical eddy size changes. To partially address the effect of the number of lags used for various heights, we have expanded the discussion in Sect. 4 using the southeast sonic anemometer data to discuss how the number of lags used changes the accuracy of the measurement at different heights, as the typical size of turbulent eddies grows larger at increasing heights.

4) Figure 10 shows the comparison of vertical velocity variances at 150 m for a) OU DL and d) LLNL WC both at 150 m with sonic anemometers. Even though closely located, large scatter is present between 0 - 0.5 m2s-2 for LLNL WC than the OU DL. Why? The differences in the R2 and the slope between the uncorrected and corrected variances are small in Fig. 10. The difference seems to be notable only at 300 m at low wind speeds. Please explain

At 150-m, there is larger scatter in the LLNL WC that is not present in the OU DL at the same height since there was an extended period of reduced SNR for the LLNL WC, whereas the OU DL maintained a higher SNR due to the fact that its laser pulse energy is larger. A sentence has been added to the manuscript to clarify why these points are scattered at this time.

The reviewer is correct in that, generally, the autocovariance technique often only modifies the value of the measured variance slightly, except under time periods when more significant noise is present or there is a significant underestimate of the variance due to spatiotemporal averaging effects. This is largely due to the fact that noise in DL data is often small, but it varies based on the SNR over time. As noted within the paper, there is significant noise present in the OU DL data at 300 m for these stable conditions. This is why the difference between uncorrected and corrected variances is most notable at 300 m for low velocity variances. We have added a few sentences to Sect. 5.3 to discuss these effects.

Specific comments:

-The authors state that Barlow et al. (2011) and Fuertes et al. (2014) already showed that DLs are incapable of resolving turbulence on small spatial and temporal scales compared to sonic anemometers due to sampling frequency discrepancies. The authors can highlight the new findings they are adding in this manuscript in more detail.

We have added more detail to the conclusions highlighting the new findings with regard to these studies.

Please also provide information on the previous studies of the variance estimates in stable conditions using auto-covariance technique if available

Several references of this have been provided on p. 8, line. 12.

-Please give the methodology used for data quality check from Doppler lidars used for the variance estimates.

Additional information about a spike filter that was used on the lidar radial velocity measurements is also provided in the description of the Doppler lidars on p. 5, lines 17-19. General information on data removed due to tower wake or precipitation is provided in the manuscript on p. 16, line 26-34.

-Page 1 line 9 :... over a variety of atmospheric conditions? Please justify the sentence

This descriptor has been removed from the updated abstract.

-Figure 1 shows a schematic of the eddies. This is already well detailed in the text lines 21-28 page no 2. Please replace this figure by showing the measurement site with instrument located.

The authors have considered this recommendation and prepared some example figures that were made to potentially replace Figure 1 (see Figs. 1 below). However, we feel that Figure 1 in the manuscript is more informative than a figure showing the measurement site with instrument locations. Figure 1 also provides a schematic of the 2 lidars and a side view next to the tower, which an overview of the site cannot provide. We also believe that the schematic shows the effect of volume averaging, which will help some readers better understand the problem. As such, we plan on keeping Figure 1 in the manuscript as is, but thank the reviewer for the suggestion that we looked into.

Fig. 1 – (left) Overview of the BAO site, with location of the tower marked with the red circle. (right) BAO tower with deployment locations of the two lidar systems shown.

-Page 12 lines 5 -10 : Please clarify the sentence- in which conditions. Please include

how many 30-min periods were available for the study for different conditions.
We are not sure which sentence this reviewer is referring to, exactly. Within the discussion of the meteorological conditions at the end of Sect. 2.3, we have added a table to indicate how many 30-min periods were available for the different conditions.

-Page no 17 line 5- 10: The authors state that differences in sampling volume and frequency have stronger effects during stable conditions. Please show a time series of vertical velocity variance comparison obtained for the 2 day period (30min, so 96 data points should be possible) at 100m and 300 m. This would be give a better visualization for understanding the diurnal variability of the vertical velocity variance comparison between the lidar and sonic measurements at different heights. Similar to Figure 2.
We have added a new figure to the manuscript (Fig. 10) showing a time series of vertical velocity variance at 100 and 300 m. We have also added a discussion of this figure.

- Please think of adding at least one or two height ranges close to the ground around 60 m in the Figure 10 to show the vertical velocity variance in the surface layer. The authors also state that the SNR was large at 50 m for LLNL WC measurements.
We would like to add another comparison closer to the ground within the surface layer. However, the data quality of the NW sonic anemometers at 50 m (and 250 m) was significantly worse than at the other heights. The recorded data had many more spikes in the various velocity fields. As such, a comparison of statistics at 50 m would require much more stringent quality control of the data and the comparison would not be as straight forward and robust. Thus, we have decided not to add another comparison closer to the ground at 50 m, the only height where there would be overlap.

- Please add a new figure comparing the variance profiles between two lidars for different conditions including the noise and sampling errors for both lidars. Also, include the sonic estimates to show the improvement of the technique.
We have taken into consideration this suggestion. However, since the SNR was often low except at the focus heights of the two different lidars during this experiment due to the clean air (as mentioned in Sect. 2.2.1, 2.2.2), there is little overlap of measurements between lidars during the experiment. Also, since the lidars and sonic anemometers are close to the tower, the tower often affects turbulence on either side as it distorts the flow. Even when neither sonic is directly waked, the TKE between sonic anemometers on either side do show some scatter with each other (see McCaffrey et al. 2016). With these issues, it is difficult to make a direct comparison between measurements on either side of the tower. As such, we have decided not to include a comparison between the measurement profiles from both lidars.

- Please include a new figure similar to figure 11 with respect to noise and sampling errors.

As discussed on p.16 l.20-25, sampling errors are small (variance errors were <5% of the calculated value for 84% of all 30 min averages) for all the data shown in this experiment.

Additionally, the experiment was designed in a way to minimize sampling errors as a possible source of difference between the lidar and sonic anemometer measured statistics. Since the sampling volumes were within a few meters of each other, which is much smaller than the range gate size for the DL, sampling errors are not expected to be a significant source of error when comparing sonic anemometer and lidar measured values. If the measurement volumes were much farther apart, sampling errors would be more important and would be more closely examined. As such, we have considered the suggestion made by the reviewer, but have decided not to include such a figure. The representiveness of the statistics is not within the focus of the paper, as we do not have the proper measurements to address this issue.

References:

McCaffrey, K., et al., Identification of tower wake distortions using sonic anemometer and lidar measurements. Atmos. Meas. Tech. Discuss., in review, doi:10.5194/amt-2016-179, 2016.