We thank the Referee #1 for the time and effort involved in reviewing our manuscript. We appreciate the very constructive and useful input a lot. See below for our point-by-point response.

The manuscript describes two approaches to reduce the effects of averaging and cross-correlation between the different velocity components that typically compromise turbulence measurements from lidars and complicate the comparability of those measurements with in-situ point measurements, e.g. performed by sonic anemometers.
The topic is highly relevant under the aspect of the potential and limitations to derive reliable atmospheric turbulence data from lidar remote sensing and falls clearly into the scope of AMT. The manuscript is clearly structured and and gives a thorough description of the theoretical background and the mathematical formulation of the models to simulate the different lidar scanning and analysis methods. It is, however, not always easy to read. The introduction would be strengthened by some discussions on the implications of the performed study on practical applications e.g. for future boundary layer research in general, and for wind energy applications, as e.g. related to the investigation of loads of wind turbines, in particular. The main weak point of the study is the very limited data set used for the comparison between measurements and simulations (only about 5 consecutive hours during one day), representing one single situation of wind conditions. In terms of wind energy, this wind speed is clearly above the rated wind speed of typical wind turbines, with the blades already considerably pitched and thus not so prone to turbulent loads. For corresponding investigations and analyzes it would be highly interesting how the method works around and below rated wind speed, e.g. for 12 and 8 m/s. Investigating more situations might also enlighten further on the large discrepancies (in the same magnitude as discussed improvements due to the new proposed processing methods) between target spectrum and sonic measurements for u and v. As long as this is not sufficiently understood, I would also doubt the significance of many of the presented results.

We agree with the referee’s helpful suggestion and have added some short discussion on the implications of improved lidar derived turbulence spectra to the introduction.
The referee considers the very limited data set we use for the comparison between measurements and simulations as the main weak point of our manuscript. We agree that an investigation with more data would have been beneficial. The reason for working with this limited data set lies in the fact that very little data are available where a commercial VAD scanning wind lidar, collocated to a meteorological mast, is scanning continuously at one height level, while saving at least the line-of-sight velocities. Currently, the only option to save line-of-sight velocities acquired by a ZephIR 300 is to stream the data manually to a connected PC. The situation gets further complicated by the fact that in the normal “profiling mode” the lidar focuses to 38m periodically for filtering purposes. So, the only known way to focus at one particular altitude continuously is to switch the unit to “turbine mode”. In this way, we acquired some data for the investigation, but their overall quality was lower than the historic data that we eventually selected as the best available data.
This data set shows turbulence spectra that strongly deviate from the shape of spectra derived from the Mann spectral tensor in the low wave number region for both horizontal velocity components. The vertical component fits well. Larsén et al. (2016) are confronted with a similar situation while they investigate spectra based on wind data of complete years for the same location, Høvsøre. The understanding of this behavior is that mesoscale effects are visible in the spectra which are not considered by the turbulence model. We can therefore not necessarily expect to reach a good fit between model spectra and sonic measurements by simply using different or more data. Instead, we based our discussion on relative spectral energy distributions

between the different data processing methods and compared them with the corresponding model predictions. In this way, we did not depend on a fit of the absolute values. This situation is unfortunate but not alarming. In particular, the effects of the two introduced methods are found mainly in the higher wave number region where the fit is at least better.

In particular, the referee raises a concern about the high mean wind speed during the measurements. Such a high mean wind speed is clearly far above average, but for the turbulence spectra it is of no importance if the wind speed is above or below rated.

Applying the method of "squeezing" to data acquired by the DBS scanning wind lidar Windcube is the object of another ongoing investigation. In that case, we have a huge amount of suitable data available and might draw conclusions with respect to different wind directions, measurement heights, atmospheric stability conditions and so on.

Here, we do not intend to investigate specific turbulence situations but want to show that the ideas behind the two novel methods work in practice and lead to the results that can be predicted by the numerical models we present. We also see and point out the limitations of the methods. Thus, we believe that the use of more or different data would not lead to different conclusions and is therefore not essential.

Larsén, X. G., Larsen, S. E., and Petersen, E. L.: Full-Scale Spectrum of Boundary-Layer Winds, Bound. Lay. Meteorol., 159, pp. 349–371, https://doi.org/10.1007/s10546-016-0129-x, 2016.

Specific comments:

1) Title: "Better" in the title immediately associates with "than what"; I would at least think about changing to "Improved", if not extending the title a bit more to make clearer what the reader can expect.

The word "Improved" was part of a preliminary title. Eventually, we decided for the word "better" as the simpler of the two synonyms. We actively decided against a more detailed title because both data processing methods are new and as yet unknown. Giving for example their names would, unfortunately, not be sufficient to explain what we did.

2) P1, L20: the actual manufacturer/distributor should be mentioned here

The name of the manufacturer "Zephir Ltd. / ZX Lidars" was added.

3) P2, L14 (and other instances): "however" should usually be separated by comma on both sides

The missing commas were added.

4) P2, L24-25: what is a "correlation twice as strong", please quantify in more detail; in addition is the fact "that the correlation increases when a time shift related to the mean wind is taken into account" presented as surprising, but it is of course not and a very standard method in many measurement applications (e.g. to correct for wind speed and direction dependent time shifts between dislocated sensors, e.g. sonic anemometers and infrared gas analyzers for the determination of latent heat and CO2 fluxes)

The increased correlation is of course not surprising, and it was indeed the motivation for the squeezing method. Although, as the referee mentions, retarding signals is used in many different applications, it is new in the field of profiling wind lidar. Bardal and Sætran (2016) give an indication of the partial validity of the assumption of frozen turbulence for wind at scales that are relevant for our work. We added some information about the setup and described the results of Bardal and Sætran (2016) in more detail.

5) P3, Figure 1: unconventional coordinate system, in particular for meteorology, where usually u defines the E-W direction and v the N-S direction

Since *u* is defined as the wind velocity in the mean wind direction, *u* can also be aligned with the E-W direction. We agree that the left-handed coordinate system is confusing when the reader expects the one that is usually used in meteorology. We want the coordinate system to be the same as in Sathe and Mann (2012) and other publications so that the model for DBS scanning given there and the one we present here for VAD scanning use the same notation and signs.

6) P4, L14-15: The quantitative results of those other experiments should be shortly presented with the corresponding references; I also miss a reference to one of the basic statements on the issue by Willis and Deardorff, 1976 (ï Ą ¸sU/U < 0.5)

That is true. We included the statement of Willis and Deardorff, 1976 and added the qualitative results of Schlipf et al., 2010.

7) P5, L17: maybe better "in contrast" instead of "by contrast"

We changed the formulation according to the referee's suggestion.

8) Section 2.4.1: would be nice to quantify and present in a table some of the key parameters for the used instrument, e.g. lR, df and a0

A reference to Table 1 that gives these values is now added in the text.

9) P6, L18-19: "... we cannot distribute the lost small scale fluctuations on spectral frequencies...", but are you able to quantify the overall amount/importance?

The formulation is now different: "The effect of line-of-sight averaging is considered in the numerical models and the discussion in this study. But none of the presented data processing methods can avoid the line-of-sight averaging effect."

10) P6, L29: insert "," after "composed of"

Thanks, a comma is inserted.

11) P8, Figure 2: some of the labels are by far too small and not or only very difficult to read; please improve

Figure 2 has been improved. We substituted the tiny label "Lidar" by a yellow symbol and updated the caption. All remaining labels are now according to the font size of the text.

12) P7, L12: "This leads to a redistribution of energy among the velocity components u, v, and w." If it is a pure redistribution this should mean that TKE spectra derived from the measurements should be unaffected, correct? Gives this a tool for a potential quality assurance of the lidar derived variances?

Good idea, but TKE spectra are not unaffected by the cross-contamination since the half cone opening angle is not equal but <45 deg. Thus, we are faced with a higher sensitivity for vertical fluctuations than for horizontal fluctuations. We do not know how much energy will be redistributed and can therefore not estimate if the TKE increases, decreases or if it stays more or less constant by the application of the squeezing method. All three cases might occur depending on the degree of anisotropy of the prevailing wind.

13) P12, Figure 4: please specify in the caption altitude level and average wind speed for the presented example

The caption now contains the information on U, D_C and f_S.

14) P12, L4: insert blank after reference Mann et al. (2010)

Done.

15) P13, L4-6: can you quantify/reflect upon the order of magnitude of the uncertainty that is introduced by this assumption

If the assumption was violated, we would need to look at all the particular beam combinations as in e.g. for the pulsed lidar modelling in Sathe and Mann (2012). This would lead to very complicated calculations. As a result, we would see cross-contamination only for resonance frequencies that are multiples of f_s. We added to the text that “It is difficult to assess the magnitude of the error committed by the assumption of continuous measurements, but we assume it is negligible.”

16) P14, Equation 23: should it read sin2c instead of sinc2 ?

Eq. 23 contains the cardinal sine function *sinc* that not everyone might be familiar with. We therefore inserted an explanation into the text.

17) P17, L18-21 (measurement description): very superficial description here, I suggest to include a few key infos in addition to the Pena et al. reference; what was the measurement frequency of the sonic; is it ensured that the measurements are unaffected by flow distortion/mast effects during the investigated case

We included more information about the measurement setup, and we have now explained why we do not expect flow distortion in the selected measurement data. The measurement frequency of the sonic anemometers was 20 Hz. This information is added to section 5.2.

18) P18, L5: “The data rate is thus for all methods 1 Hz”; does this also include the sonic data? If so, how are the sonic 1 Hz data sampled? I assume the sonics run somewhat between 10 and 50 Hz, so you could create 1 Hz data either by averaging your raw data, or picking one raw value every s. The corresponding selection will of course have an influence on your final spectra.

Good point. The following information is added: "The sonic anemometer measures with a rate of 20 Hz. This high frequency data is down-sampled by the use of the MATLAB function 'resample' to a frequency of 1 Hz. The function includes a low-pass filter to avoid anti-aliasing."

19) P19, Figure 5 (also Figures 6 and 7): I suggest to reconsider your presentation of the results ; in particular the black and blue squares are very hard to distinguish; this makes it very hard to follow the discussion of the results; a quick fix could be to consider in addition to the different colors also different symbols (e.g. square, star, triangle)

We appreciate the insight that some of the results might be difficult to read. For a better distinction of the results presented in the plots, we have replaced the square markers by triangles with different orientations. We also replaced blue lines and markers by cyan colored ones.

20) P19, Figure 5: why is the target model so far away from the sonic measurements; are there any effects of flow distortion visible (connects to comment 17)

Analyzing the definite cause for the discrepancy between model and reference lies unfortunately out of the scope of this study, but we have extended our explanation of it and added a second, better reference: "The extra energy at low wave numbers compared to the spectral tensor model for this site has been observed before and is related to the inhomogeneous landscape at Høvsøre with its sea to land transition in the main wind direction (Sathe et al., 2015) and mesoscale effects that overlay the expected spectral gap (Larsén et al., 2016)."

21) P21, L16: "... show very good agreement to the model"; I feel this is a very strong statement; I can support this only for kl>10-2, but not for the rest

We changed our statement to: "Overall, spectra calculated from the 2-beam processed measurement data show good agreement to the model." And added the following: "It is important to keep in mind that, due to the poor fit of the measured spectra of the horizontal wind components and the modeled spectra at low wave numbers, we can compare the relations between the different methods but not absolute values."

22) P23, L14-16: the unfortunate mis-representation of the sonic measurements by the target spectrum is again a great concern; how confident can you be in the discussion of improvement caused by your new processing methods, when the discrepancies between the target model and the sonic measurements are of the same magnitude or even larger; Again I would strongly recommend to also look at and present results for different synoptic/wind situations to provide some evidence where this mis-match between target spectrum and measurements originates from.

We are well aware that our statements about effects we see in the spectra have a thin data basis compared to other investigations. And we appreciate getting the feedback that this creates doubt about the overall results. To respond we would like to point out that our analysis is also based on the theoretical considerations we elaborated in sections 2 and 3. Most of the results are as expected and described as thoroughly as possible. Only minor effects cannot be explained satisfactorily, which is then mentioned in the text. The good coherence between what we know about the theory, the model predictions and the measurement results gives us confidence in the points we raise in our discussion.

23) P 24, Figure 7: its puzzling that the sonic measurements and the target model fit very well in the case of w, while the match it is rather poor for u and v;

An explanation for the better fit of the model spectrum to the measurements in *w* than in *u* and *v* is given in Larsén et al. (2016). We added: "The results of Larsén et al. (2016) show that the spectra for vertical fluctuations are not prone to contributions from the mesoscale spectrum."

24) P26, L6-7; This is a crucial issue of the manuscript and has to be elaborated in more detail (see also my corresponding concern in the general comment section)

See our comment in the general section prior to point 1).

25) Formatting of references is inconsistent a. Journal names abbreviated/not abbreviated b. DOI given or not c. Incomplete, pages missing: e.g. Newman et al., 2016; Pena et al., 2015; Sathe and Mann, 2012;

We thank the referee for mentioning the inconsistency in the list of references. The formatting is consistent now: journal names abbreviated, DOIs given wherever available and pages added wherever applicable.