Dear Anonymous Referee #1,

Thank you for your comments. We provide our rebuttal to the points raised. Below are the author rebuttals in red and anonymous referee #1 comments in blue.

The authors present field data from a wind turbine with and without retrofitted Gurney flaps near the blade tips. The experimental data is velocity data obtained through LIDAR. One section that deals with a power and load analysis based on blade element momentum simulations is added at the end.
We want to kindly mention that the manuscript also contains some more topics which are relevant for a better interpretation of the results and conclusions. The following list of points could be acknowledged for a better summary to the readers of the review:
1. The use of Gaussian process regression in the post processing chain of scanning LiDAR measurements.
2. The analysis of different wind speed retrieval methods.
3. The atmospheric stability implications made.
4. The results of free vortex wake simulations as a validation technique.

I find the paper overall difficult to follow and not well written. The paper is full of grammatical and spelling errors.
We regret that and we have taken the point into account. We will improve the readability in the revised manuscript.

Figures are not cited in order of appearance. Variables in equations are not introduced. References are cited as a list without summarizing their main contributions and relevance to the current manuscript. Figures lack axis tick labels.
Figure citing will be updated in the revised version, as well as introduction of variables used. To the author's interpretation, the relevance of the cited research is clearly defined in the introduction.

Only Figures 1 and 6 miss axis tick labels. This is because of confidentiality reasons as highlighted in the 'Code and Data Availability section'. We want to mention that we have properly labelled all the other figures in the manuscript.

The term 'segmented Gurney flaps' is used before explaining what it means, and then on page 3, the authors suddenly start to talk about miniature trailing edge effectors instead.
We agree that the segmented aspect is not clearly defined early on in the manuscript, and will provide more clarity in the revised manuscript. As the term miniature trailing edge effectors (used in cited literature) is perceived as confusing, we will replace it with segmented Gurney flaps in the revision.

The results themselves are not convincing. The authors themselves seem to be torn. They go from admitting that the impact of Gurney flaps remains within the measurement uncertainty band in the abstract to claiming that great insights into the wind turbine wake were obtained in various conditions.

Unfortunately, this point is a serious misinterpretation of our statements. We talk about the results being within the measurement uncertainty for the 'power and loads' analysis, while the great insights remark is made for the 'wake analysis' conducted in this study. The coupling of these two leads to a major misinterpretation of our results and conclusions.

In my opinion, the data shown does not allow any conclusion to be drawn. Too much information is missing.

Could you let us know what information is missing? Additionally, the 'Code and Data Availability' section states that wake analysis data is available upon request. Also note that to keep the manuscript concise we try to only include the most relevant information for repeatability. The thesis report (cited on line 125) provides more background information.

The authors refer to the standard deviation of the data as the standard error. A distinction between data fluctuations and measurement errors is not made.

Could you further elaborate on the point? We have clearly defined the standard error differently to standard deviation, on line 229.

As evident by various existing literature sources, which are also cited in the paper, standard error is the metric which is typically used for scanning LiDAR measurements; which is implemented in the results of this study as well.

The wake deficit is presented in the form of a spanwise average. I do not understand why the authors do this, as the average value depends strongly on the spatial sampling distribution and the area over which they average. There is too much bias.

Please note that both local (50% span) profiles (vertical) and spanwise average profiles (axial) are provided in the manuscript. The spanwise average metric is also explained in the manuscript (lines 211 to 216), including how the area is chosen.

By introducing the various checks in the data set, relating to inflow conditions and wind turbine performance (see also lines 170 to 173), the comparison is made unbiased to the best extent possible. In addition to that the standard error is clearly indicated in the plots.

The authors hope that these points can be taken into account as well for a second review.