Wind Energ. Sci. Discuss., author comment AC1
https://doi.org/10.5194/wes-2021-90-AC1, 2021

Reply on RC1

Maria Krutova et al.

Author comment on "Development of an automatic thresholding method for wake meandering studies and its application to the data set from scanning wind lidar" by Maria Krutova et al., Wind Energ. Sci. Discuss., https://doi.org/10.5194/wes-2021-90-AC1, 2021

Thank you for your comment. It was very helpful for the revision of the article structure and clarify the parts that are not very common in wind energy studies.

RC1: It seems that there are only one or two examples shown for output from the ATS method using LES. I understand the LES simulation is used to test the method on an idealized case but unless you're working with a larger subset of LES scans it doesn't seem relevant to put LES in the title of the manuscript? I also wonder if you even really need the LES simulation for this study since it just seems to be used to explain the methodology.

The LES simulation serves as a simplified example to demonstrate the core methodology. We admit that the title can be misleading. The title is changed, and LES-related examples are moved into a stand-alone section 'Proof of concept' to separate them from the results. The LES examples now focus on the instantaneous wake instead of the averaged case.

RC1: The LES scan case highlights an interesting example – the ATS algorithm results in many wake shapes (noise) in the far wake; when comparing the ATS v. threshold deficit methods, are you considering these far wake shapes in the confusion matrices? It's hard to tell if you completely ignored the far wake in the analysis or if it's included but acknowledged as a source of uncertainty? Because it seems like these shapes would affect calculation of the wake centerline when comparing ATS centerline to Gaussian centerlines?

The far wake was always considered in the threshold verification. However, the current centerline detection algorithm processes only the first continuous wake shape, so the centerline detection from the ATS-detected wakes stops for some scans while the Gaussian method can continue (that does not indicate that it will succeed, though). We acknowledge the current limitation of the ATS centerline search, but plan to improve it further.

RC1: The manuscript could benefit from better figure organization (some of the figures appear before they are called or are shown in incorrect sections, more on this in 8.). It seems the equations in section 4.1 might be somewhat out of order; it's unclear where WP is actually used in the process of the ATS method?

The figure and table positions are mainly caused by the LaTeX formatting. Since the corrections affecting the paragraph length could change a figure position, we left it for the final editing.

WP matrix is now included in the algorithm description as a value.

RC1: It would be interesting (and likely important) to know how the radial velocity values are transformed from radial to cartesian components.

This could be a confusion caused by wording. The lidar data used in the thresholding are always stored in the rectangular matrix defined by the polar coordinate system (beam range, azimuth). This matrix is processed both by the thresholding and centerline algorithms. However, if plotted as it is, the polar matrix distorts the scanning filed in the far and near lidar range. Hence, we converted the coordinate field from (beam range, azimuth) to (x,y) to plot a scanned sector in a more human-readable way.

The thresholding method always used the radial velocity, because this method does not strongly depend on the actual wind speed values, but mostly their relative value.

RC1: The word "threshold" is used many times throughout the manuscript (i.e. in the introduction) before it's completely clear what type of threshold is really being applied; it would help to clarify this for a broader scope of readers.

The threshold is now explained in the introduction when it is first mentioned:

In a simplest case, a threshold is a value that splits the range into two parts: all values below the threshold fall into one group, while the values higher than the threshold form the second group. When applied to the wind field for the wake identification, a threshold would split the data into the wake and free-flow points.

RC1: It might be helpful at the end (or beginning) of the methodology to have a table outlining and summarizing all of the techniques you've used, the amount of data (and what type of data) you've used for each technique, and to also clearly show which techniques you are comparing. You're using two common techniques for comparison, comparing manually detected shapes and ceterlines, and also switching up the datasets for some of them (i.e. the inclusion of the LES case study) so it is rather hard to follow. In the first paragraph of the Results section you broadly summarize your process but I don't think it's enough to help with the organization. Maybe even call these comparisons Experiment A, B, etc. and label that in the table and section headers? Just for ease of organization.

Thank you for the suggestion. A summary table is added to the end of the Methodology section listing the method names, properties and the input data.

RC1: Do you have ensemble statistics for comparing Gaussian and ATS methods for centerline detection? As I understand it, Figures 16-19 are interesting case studies but the overall success of ATS v Gaussian is still ambiguous. Maybe you could have only one figure showing the R^2s for each wake (because it's helpful for visualization) and then have a table for ensemble statistics (so you would reduce your figure count by about 3).

The direct comparison between Gaussian and ATS method is rather complicated, because the Gaussian method does not perform well enough on the noisy scans and merged wakes to be considered a good reference. R^2 statistics worked when the manual centerline was presented as a 'true' value. We planned a comparison based on visual inspection: score the success rate of the Gaussian and ATS methods in the near and far wake. This comparison was scrapped in favor of the verification against manual detection, but we will consider to include it back.

RC1: Figures, figure labels and captions should be clarified:

RC1: Figure 4 could use a label for the intensity colorbar below, and the caption states a "valid wind speed range" but this range isn't clear numerically.

RC1: Figure 5 could use a label for the intensity colorbar below and labels for the entropy figures (a) and (c). It might also be helpful to put the location of turbines in this image as you did in other figures farther below.

RC1: Figures 7 and 8 and Table 1 should be in the previous section.

RC1: Table 2 should be in the previous section

All corrected

RC1: Labels for Figure 10c are ambiguous – what are helper lines?

This was an error in the plotting script, the 'intersection' labels should refer to the black dots. 'Helper lines' are labelled correctly and refer to the concentric circles described in the centerline detection algorithm.

RC1: Figure 11 is quite hard to follow in general, and it's the first time the colorbar has been used to distinguish different wake shapes. This is another instance where discussing the output of the ATS (in terms of producing multiple wake shapes) would be helpful. Do you need the black dotted centerline in (d)? It might be helpful to switch (c) and (d) and remove the black centerline in d, also maintain a red color outline for the wake in (c) and (d) so readers don't think you are using different methods to get the red and black wake outlines.

The figure is changed to show different methods in two different subplots + overlaid wake contours and centerlines in the third subplot. This allowed to keep the same legend without color conflict. The color-coded wake shape are now moved to Fig. 10.

RC1: Figure 15 – I'm not sure why the corrupted scans are included in this figure since they were difficult to inspect manually?

Some of the corrupted scans allowed to perform a manual detection. We agree, that it adds unnecessary complexity to the plot. Due to the corrections in methodology for the deficit-based threshold, the thresholds cannot be compared directly anymore. This plot became obsolete and is replaced with a box plot summarizing the percentage of true positives and true negatives for different methods against manual thresholding.

Figures 9-11 and 15 had experienced the most changes, so they are attached as a supplement.

RC1: Figure 19 – is the "missing" gray color referring to the non-filled gaps?

The 'missing' label is replaced with 'no data' in all relevant plots to improve clarity.

RC1: Figure 20 – not sure what invalid/valid and WIND and WAKE=WIND mean? Could you clarify?

The labels were dictated by the brevity to leave more space for the actual plot. The label for the scatter plot are changed as following:

'invalid' ☐ 'corrupted'

'valid' ☐ 'data'

WIND and WAKE refer to the respective directions and are labelled near the respective axes.

RC1: The conclusion feels rushed and some of the statements in the conclusion about success rates of the Gaussian v. ATS in the near wake seem confusing / misleading – are you using ATS to identify centerline for far wake objects? Again it'd be helpful to have ensemble statistics for Gaussian v. ATS and better clarification as to whether or not far wake objects are included in success rate analysis.

The far wake centerline is detected by the ATS method, when the detected wake shape is long enough. If the detected shape ends at, e.g., 4D, the ATS method cannot proceed, while the Gaussian method can – although it does not guarantee valid results for the Gaussian method. See the comment about ATS to Gaussian comparison.

The conclusion is now re-written to be more detailed on the methods used, their level of automation, advantages and disadvantages.

Minor Comments

Line 174 The word "significant" shouldn't be used unless significance testing has been done

Removed

Line 125 The reference wind direction is mentioned several times and is an important part of the manuscript but it's unclear whether this is wind direction as given by the met mast?

The reference wind and direction were mentioned in lines 98-99. We agree, that the mention was too brief and had to be more emphasized. Section 2 is now rearranged to describe the site, reference data and lidar scans in enough details.

Line 141 Why did you choose the wind speed of 7 ms-1 as appropriate?

The line should be read as 'wind speed difference'. The value is chosen empirically for this particular data set.

Line 105 It's a bit hard to follow when the reader has to scroll many pages down to Table 1; is it worth not mentioning table 1 and the scan subsets in this

sentence instead mentioning them later, possibly in the data quality section.

Mention is removed

Line 163 Directional entropy is a bit of a misleading term. I was thinking it would be entropy in wind direction measurements of some sort; is there a better term to use for this?

We chose this term to have a short reference to the Shannon entropy calculated across the beam range or azimuth. Apparently, it matched a pre-existing term, which definition is different from what we had implied. The 'directional entropy' is now replaced by the 'row and column entropy' to avoid confusion.

Lines 163-185 This discussion would benefit from including numerical entropy values for the scans discussed, i.e. "The entropy calculated across the beam range highlights several lidar scans with an entropy decrease compared to other cases (insert mean entropy values for these cases)" – as is, the point of this paragraph is somewhat hard to follow although the figure is helpful; could we have a succinct sentence summarizing these results at the end of the paragraph marked by line 187? It would be helpful to include numerical values to further contextualize Table 1 as well. Also it would be important to clarify what constitutes "corrupted" measurements – is this based on SNR/CNR filtering? Or is it based on the number of data points that exceed the wind speed limit as you discuss later? Either way, it would be helpful to clarify what this means the first time it's mentioned.

The entropy values are added.

The corrupted measurements refer to the non-physical values of the wind speed (100-1000 m/s) appearing in the scans due to the crosswind effects mentioned earlier in lines 115-130. The mention of the corrupted scans is now added to the line 130. The entropy criteria are now described in more detail.

Line 189 Can you be clearer about the numerical value of the percentage of data points that exceed the wind speed limit? Since this is an important criterion for identifying "corrupted scans"?

Added the following lines:

The corrupted scans in the data set consistently have at least 1% of points exceeding [30 m/s] limit. The percentage drops to 0-0.05% for the rest scans and corresponds to the occasional spikes.

Line 190 Why are the corrupted scans split into 5 subsets / why is that relevant?

The data set was initially split into subsets. The corrupted scans cover exactly five subsets and do not appear in other subsets. The total number of the corrupted scans is more relevant, indeed, so we will remove the mention of the subsets. Their count would be still seen from the overview table.

Lines 195-214 These should refer to Figure 8, not Figure 7, right?

Corrected

Lines 220-221 Can we have more detail about the phrase "shown to reproduce the wake shape rather accurately"?

Details about PALM reproducing double wake and polynomial kernel of the model are added.

Line 261 I think it's worth stating in the previous paragraph that the output of the thresholding will result in multiple different shapes (particularly in the far wake), since this isn't visible for the initial LES case (since it's continuous). It's an important result of the thresholding algorithm so I think it's important to discuss it thoroughly and to discuss it well before you show it in Figure 11, for example.

The LES section is now re-organized to demonstrate the wake detection on an instantaneous wake.

Lines 324-326 You could probably remove this paragraph.

Removed

Lines 143-144 How many points constitute a "larger cluster" or high wind speed values? So if gap filling is not performed, are there just missing data values within the scan? And these are thus visible in the scan image?

We performed gap filling only for 1-3 adjacent spikes removed, as they left enough points around to perform a simple interpolation without introducing uncertainty to the data. The whole preprocessing procedure was aiming to clean the lidar scans from the obvious outliers, but not alter the original data too much.

The gap filling was not performed for the deleted points (values above 30 m/s). The thresholded images would show those points as 'missing' ('no data' in new notation). No distinction is made between originally missing data and values removed during the preprocessing.

Line 330 I'm not sure what "The deficit-based threshold is presented as the wind-speed threshold" means – can you clarify?

This is the threshold calculated as 95% of the free flow wind speed (Uref). It does not require for the wind field to be scaled to [0,1] range, like ATS method, but is applied to the velocity values. Since we had found an error in the method implementation to the radial velocity, the paragraph is removed and the whole subsection on the deficit-based thresholding is re-written.

Line 347 How many scans do you choose to demonstrate the methods' performance? Why do you choose them? Again might be helpful to put this in a table.

We chose an example from each of the most represented non-corrupted subsets (parallel, aligned, bimodal) to demonstrate the method performance in each case and highlight the potential problems. Where possible, we tried to stick to ensemble statistics, since the scan selection could be highly subjective. A supplementary video was provided to present results of the wake identification for the whole data set.

Section 5.2 title should include a mention of the wake deficit method since you're also comparing that with manual detection and ATS

The subsection title is changed to 'Comparison of the ATS wake identification against the manual identification and deficit-based thresholding'.

Section 5.2 Would it be helpful to refer to manual detection as "ground truth" – so we know that you are comparing the performance of ATS and wake deficit to that of the manual detection and producing the confusion matrices as such? This just goes back to ease of organization; maybe a table as mentioned previously would make this clearer.

The manual detection is now emphasized as a 'true' detection.

Table 2 To understand the results better, could you explain why the manual-deficit exhibits such a comparatively high rate of false positives?

Due to the improved implementation of the deficit-based threshold, the table will be different now. The remark is still relevant for the selected subset, although the difference is not so dramatic. Since the deficit-based wake identification defines a threshold through a fixed relationship 0.95Uref, the method does not perform self-adjustments depending on the data quality. It appeared, that the deficit-based threshold is underestimated compared to the manual thresholding and identifies more points as wakes while they actually belong to the background flow (false positives).

Please also note the supplement to this comment:
https://wes.copernicus.org/preprints/wes-2021-90/wes-2021-90-AC1-supplement.zip