Wind Energ. Sci. Discuss., referee comment RC1
https://doi.org/10.5194/wes-2021-90-RC1, 2021

Comment on wes-2021-90

Anonymous Referee #1

Referee 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-RC1, 2021

The work presents an interesting, new technique for wake identification and characterization from lidar scans using principles from image processing. The work presents promising results for wake characterization using a novel image thresholding method (applied to another use case in a separate field and used for the first time for wake characterization in this manuscript). Compared to existing velocity deficit thresholding techniques, this method exhibits a higher success rate for wake characterization. The paper is interesting and novel and considers the use of a LES simulation case study to explain the novel thresholding technique and compares multiple methods for wake characterization (Gaussian, velocity deficit threshold, the novel thresholding technique, manual inspection). However, these comparisons are hard to follow and the paper could benefit from a better organizational structure and improved figures. Ensemble statistics for successful wake characterization are only given when comparing the new thresholding technique and the velocity deficit thresholding technique previously used in the field. The paper could benefit from further clarification as to whether they are considering the far wake in their analysis and from displaying ensemble statistics comparing the novel thresholding technique to the Gaussian technique (it seems that these statistics are only shown for case studies within the scan subsets in Figs 16-19; if this isn’t the case then clarification is needed). Since only lidar scans are used to compare success in wake characterization between methods, it's questionable if the LES case study is really adding anything to the manuscript.

Major Comments:

- 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 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 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?
- It would be interesting (and likely important) to know how the radial velocity values are transformed from radial to cartesian components.
- 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.
- 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.
- 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).
- Figures, figure labels and captions should be clarified:

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.

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.

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

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

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.

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

Table 2 should be in the previous section

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

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

- 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.

Minor Comments

- **Line 174** The word "significant" shouldn't be used unless significance testing has been done
- **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?
- **Line 141** Why did you choose the wind speed of 7 ms^{-1} as appropriate?

- **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.
- **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?
- **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.
- **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"?
- **Line 190** Why are the corrupted scans split into 5 subsets / why is that relevant?
- **Lines 195-214** These should refer to Figure 8, not Figure 7, right?
- **Lines 220-221** Can we have more detail about the phrase "shown to reproduce the wake shape rather accurately"?
- **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.
- **Lines 324-326** You could probably remove this paragraph.
- **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?
- **Line 330** I'm not sure what "The deficit-based threshold is presented as the wind-speed threshold" means – can you clarify?
- **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.
- **Section 5.2** title should include a mention of the wake deficit method since you're also comparing that with manual detection and ATS
- **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.
- **Table 2** To understand the results better, could you explain why the manual-deficit exhibits such a comparatively high rate of false positives?

Please also note the supplement to this comment:
https://wes.copernicus.org/preprints/wes-2021-90/wes-2021-90-RC1-supplement.pdf