Reviewer comments

Thomas Duc, Arthur Finez

June 25, 2026

1 Response to anonymous referee 1

This manuscript investigates the effect of yaw misalignment on wind turbine noise emissions through a full-scale acoustic measurement campaign. The experimental design addresses a relevant knowledge gap: quantifying acoustic implications of wake steering strategies increasingly deployed for power optimization. The study reports statistically significant directivity asymmetries that challenge omnidirectional noise source assumptions in current acoustic models, and observes a modest but consistent increase of approximately 0.6 dB(A) in overall sound power level under yaw misalignment. While the experimental approach is innovative and the dataset appears comprehensive, the manuscript suffers from insufficient justification of key methodological choices, inadequate discussion of contradictions with prior literature, and presentation issues that limit accessibility of the results.

The work represents a valuable contribution to the WFFC literature and provides high-quality data for model validation. The experimental design and dataset quality are strong, but methodological transparency, contradiction with prior literature, and presentation quality should be improved before publication. The authors should provide justification for processing parameter choices and discussion reconciling the current results in the context of past work in wind turbine aeroacoustics and directivity.

First, thank you very much for your positive feedback and the comprehensive review of the paper. We tried to account as much as possible to your suggestions in the revised version of the paper, as you can read in our comments below. We hope that these additions contribute to strengthen the scientific content of the paper.

1.1 Major Concerns

1. The observed increase in sound power under yaw misalignment appears to contradict past literature cited by the authors, which reported noise reductions or no significant effect from yaw misalignment. The authors cite these studies but offer no substantive discussion of why the current results diverge. Are the differences from turbine model, blade aerodynamics, study or measurement methodology?
A section has been added at the end of the paper to discuss results reported in the literature and propose some interpretation for the discrepancies. In addition to this section, rotor speed curves in Figure1 show that the curve with non-zero yaw are not generally below the $\beta = 0^\circ$ case (except Yaw +20° at 8 and 9 m/s). This is a possible reason for discrepancies with the main reference of NREL.

2. Time shifts of up to ±15 seconds are substantial relative to atmospheric turbulence timescales and microphone sampling frequency. The manuscript does not explain the root cause of such large clock drift over a four-day campaign. The cross-correlation procedure reduces uncertainty to approximately 1 second, but consequences for 10-second aggregated acoustic measurements remain unquantified. Given that acoustic propagation times span 1 second across the measurement domain, this uncertainty may not be negligible for directivity pattern analysis. Can the authors speculate on the impacts of the time delays as a source of measurement uncertainty? How does this influence the final results?
We run a sensitivity analysis on this effect by recalculating all results with an ±1 second offset on the acoustic data. As can be seen on Figure 2 below, the impact of this shift is negligible on the calculation of the directivity patterns and does not change the main conclusion of the paper. A

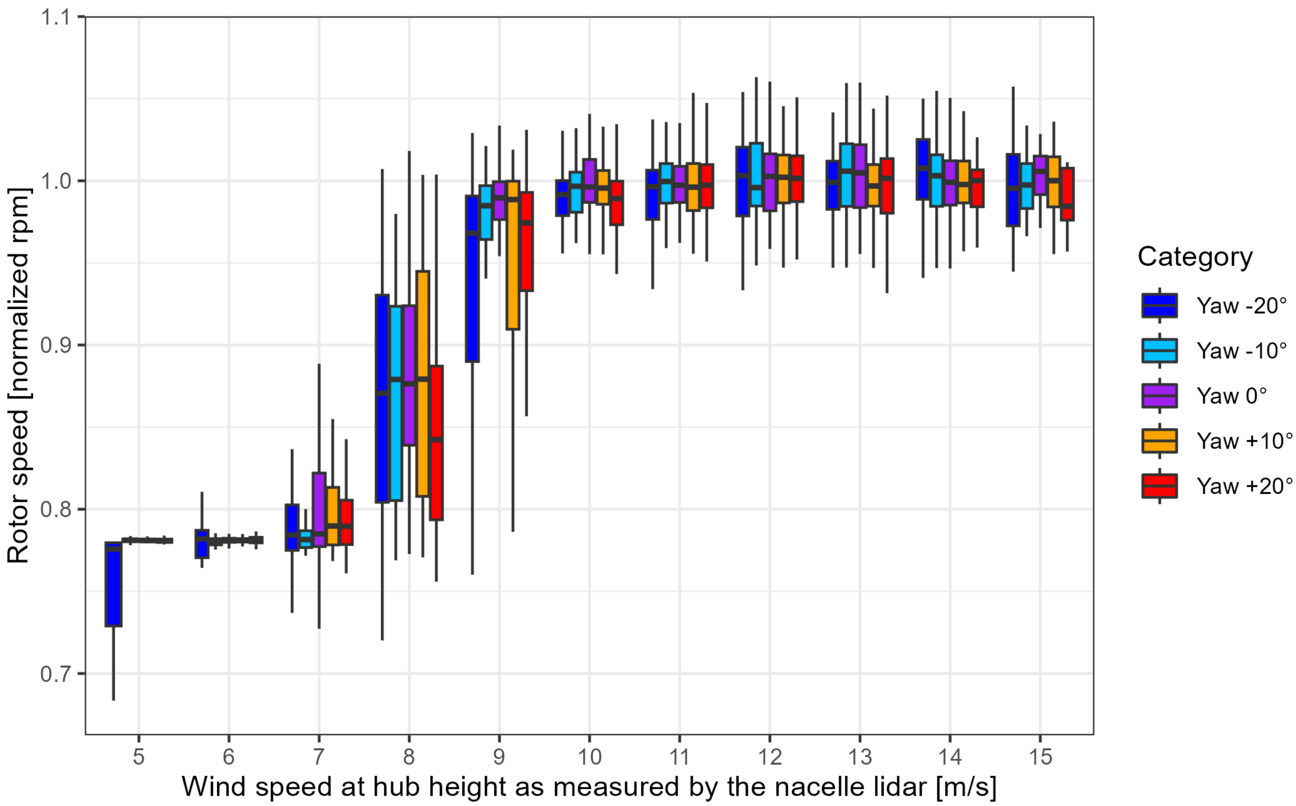

Figure 1: Rotation speed for different yaw angles. The rotor speed is normalized for confidentiality, as required by the turbine manufacturer.

sentence has been added in the manuscript to clarify this point.

3. Multiple filtering and detection parameters appear arbitrary without justification. The ±50 second rolling median window and 10 dB exceedance threshold for outlier detection (Section 3.1.2) lack supporting rationale. What is the false positive rate? How many genuine turbine noise events are erroneously filtered? Have the authors checked the manually recorded noise events against the automated event detection? The reduction of minimum samples per bin from 10 (IEC standard) to 3 (Section 3.3.2) to avoid discarding 45% of data is concerning.
Details of the parameters choice have been added in the paper. In addition, a consistency check was conducted using the field observer's notes. The observer reported 91 events over the full campaign, of which 74% resulted in a clear increase in the OASPL time series (greater than 4 dB) for SLM#1. Among these events, 54% were correctly identified by the median filter. However this test is considered of limited relevance. On the one hand, the observer may have missed some external noise events; on the other hand, some visually detected events may not be audible across all microphones and may not qualify as actual acoustic outliers. Moreover, Figures 5 and 6 of this document show that the whole procedure is relatively stable with respect to the outlier filter parameters. Regarding the reduction of the minimum number of samples per bin from 10 to 3, this is only a temporary adjustment. It is subsequently compensated for by imposing a minimum of 10 **cumulative** ***samples across microphones, per bin, relative positions, and yaw angles. A sentence has been added in section 3.3.2.***

4. The frequency-dependent nature of directivity changes (Figure 18, low-frequency effects dominate) and non-monotonic changes with wind speed (Figure 19, moderate yaw offsets produce larger changes than extreme offsets at high wind speeds) suggest complex underlying aeroacoustic physics that are not explained in the manuscript. What mechanisms separate low and high-frequency noise emission responses to yaw misalignment? Why do moderate yaw angles produce larger acoustic changes at high wind speeds? These patterns may reveal fundamental insights into trailing-edge noise, turbulent inflow interactions, or blade loading asymmetries that should be discussed at least.
A discussion has been added in section 4.2 concerning the frequency-dependent nature of directivity changes, in particular the possible influence of the yaw angle on turbulence inflow noise, whereas it is virtually ignored by turbulent boundary laryer-trailing edge noise. Concerning non-monotonic changes with wind speed, a new paired t-test was run to evaluate the hypothesis of high speed sound power levels at $\beta = -10^\circ$ ***being higher than*** $\beta = -20^\circ$. ***The corresponding p-value being above 0.5, the non-monotonicity cannot be inferred directly from the measurements and we preferred not to interpret a statement that cannot be statistically supported.***

1.2 Detailed Comments

Section 2.1, line 95: Was any testing conducted to quantify the acoustic impact of uncut versus cut wheat? Background noise measurements (Figure 11) show up to 4 dB variation across SLMs at fixed wind speed, attributed to "local environment" differences. Quantifying the wheat crop contribution would strengthen confidence that site-specific background variations are adequately characterized rather than introducing systematic bias in turbine noise extraction.
Unfortunately, we could not conduct some tests to quantify the acoustic impact of uncut versus cut wheat. This is why we decided in the methodology to calculate a background noise for every SLM to make sure the local specifities around each of them are accounted for. We added some text in the manuscript to discuss a bit more the variations of background noise for a given wind speed.

Figures 2, 3, 6, 7: SLM identifiers should be consistent across all figures and text. Adopt a single convention (preferably the full identifier for traceability) and apply uniformly. Cross-check all figures to ensure labels match the site map.Axis labels, annotations, and colorbar text are difficult to read in several figures.
As suggested by second reviewer, we finally opted for the convention "#01". This makes it easier to read in the text and the figures. All figures were re-edited with this new convention, and we also increased the font in most of them to make them easier to read.

Section 3.2, lines 220-237: The hybrid approach introduces potential discontinuities. Report the fraction of data

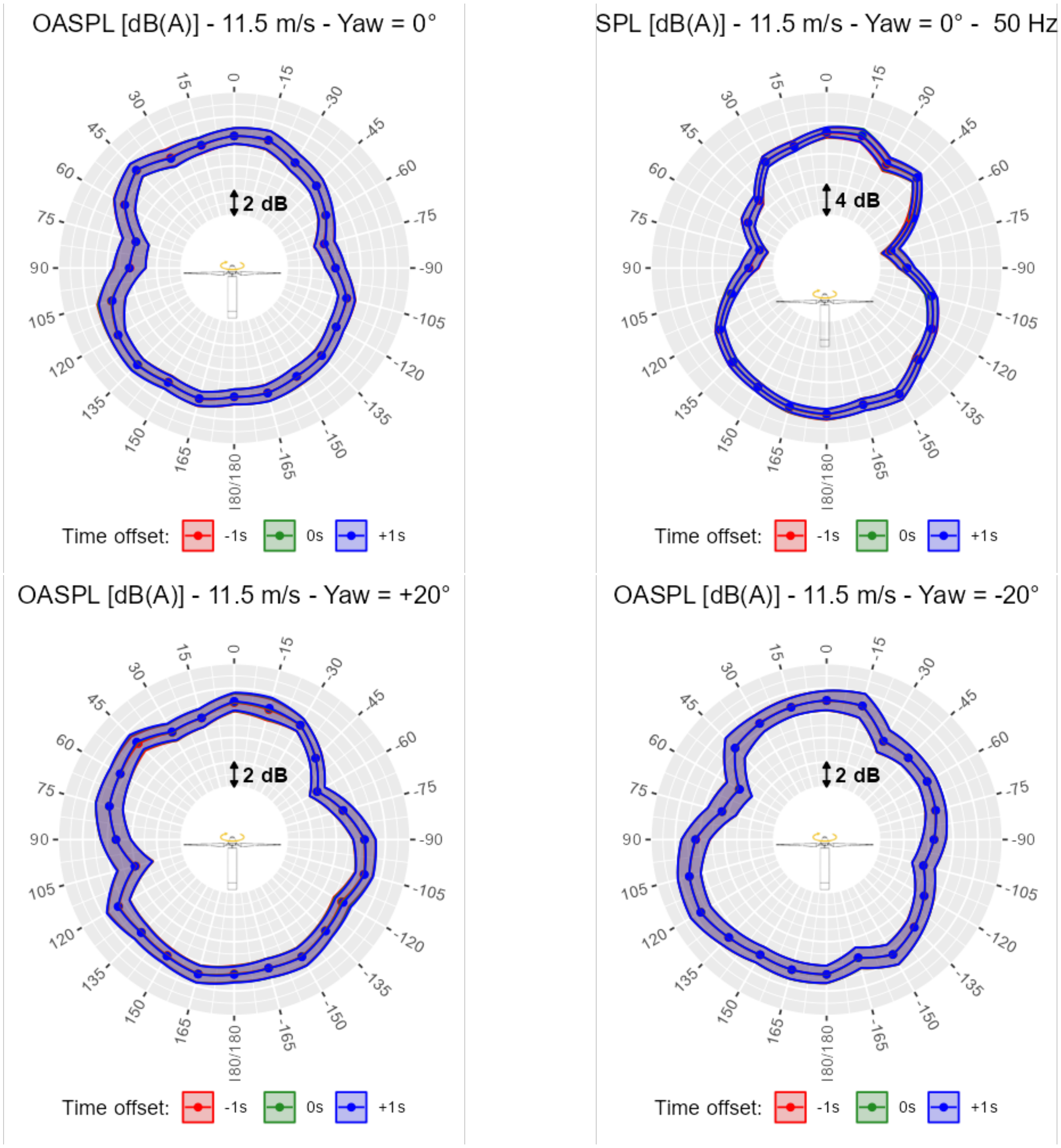

Figure 2: Results of the time offset sensitivity analysis on the directivity patterns calculated in the paper. Although only a sample of graphs are shown here, the same behavior was observed for all of them.

falling into each category and assess whether transitions between wind speed sources introduce artifacts. The 1-minute rolling average applied to nacelle lidar data may smooth out turbulence-driven acoustic fluctuations; discuss whether this temporal filtering impacts correlation between inflow conditions and noise metrics.
With the choice realized for selecting the source for the reference wind speed signal, the risk of discontinuities is much reduced. Indeed, except for the case of turbine operation under no yaw misalignment, the nacelle lidar signal is always chosen since the other sources are considered unreliable. For the 0° yaw case, there are some transitions between the inverted power curve and the nacelle anemometer sources from time to time (this happens when the turbine reaches its nominal power and the power curve cannot be inverted), but those changes are the ones that are recommended by the IEC noise standard. Finally, since we discard 3 minutes of transient data when toggling between different yaw categories, this leaves plenty of time to switch from the lidar to one of the SCADA wind speed sources. Given that the lidar is used for all yaw categories (except yaw = 0°), it is the source that is used approximately 85% of the time, while the inverted power curve and the nacelle anemometer are used only 9% and 6% of the time, respectively. These numbers and a clarification about this topic were added to the manuscript.

In order to define the best methodology for the nacelle lidar wind speed reference, we compared the correlation of different resulting signals with the reference signal coming from the inverted active power signal at yaw = 0°. This result is now presented in the Table 2 of the paper, and discussed in an additional paragraph. As can be seen, the 1-minute rolling signal is by far the one with the best correlation and show close performance with the nacelle anemometer. We therefore think that this choice limits issues in the correlation of noise metrics with inflow conditions.

Figure 16: Consider combining separate panels into a single multi-panel figure with consistent axis ranges for direct comparison. Absolute axis limits would be extremely helpful in interpreting results. If permissible, please include.
Unfortunately, we did not get the permission to show results with absolute axis limits, and we were forced to plot all graphs with relative axes only. However for this figure the two y-axes have consistent axis ranges: this can be seen thanks to the -180° curve that is common and absolutely identical in both graphs. A sentence has been added in the legend of the figure to stress this point. We prefer to keep two different sub-figures, because it makes the graphs clearer and better support the discussion in the text in which we compare the position 45°/135° and -90°/+90°, respectively.

Figure 18: Changes in directivity induced by yawed operation may be more evident if results were presented in the rotor reference frame rather than the wind reference frame. Have the authors explored this alternative representation? The strong frequency dependence (low frequencies affected, high frequencies unaffected) suggests distinct noise generation mechanisms. What aeroacoustic processes explain the frequency threshold separating responsive and non-responsive regimes? Discuss whether low-frequency changes correlate with blade loading asymmetries, inflow turbulence modulation, or wake-induced effects.
We investigated this alternative representation (see Figure 3), however the changes in directivity are not very obvious and we think it is more interesting to keep the curent representation in the wind direction frame to show the rotation of the pattern with the rotor. Also, the representation in the rotor reference frame is not ideal with the current setup, because since the yaw offsets steps are 10°, and the angular position of microphones is 15°, that forces us to resample data into 15° bins in order for the directivity patterns to be comparable.

A discussion on the effect of yaw on different aeroacoustic processes has been added in answer to question 4 above, and their relation with spectral directivity patterns. Concerning wake induced effect, we added a sentence suggesting that the observed noise increase of TI may be compensated in WFFC strategies by a reduction of turbine/wakes interactions. However, as the single test turbine was operating in free field (without wakes from upstream turbines), we do not have new material to quantify this counteracting effect. Concerning blade loading asymmetries under yaw, the blade were not equipped load sensors. Since this effect is known to be model dependent, it is difficult to correlate unknown blade loading with the noise variations we observed.

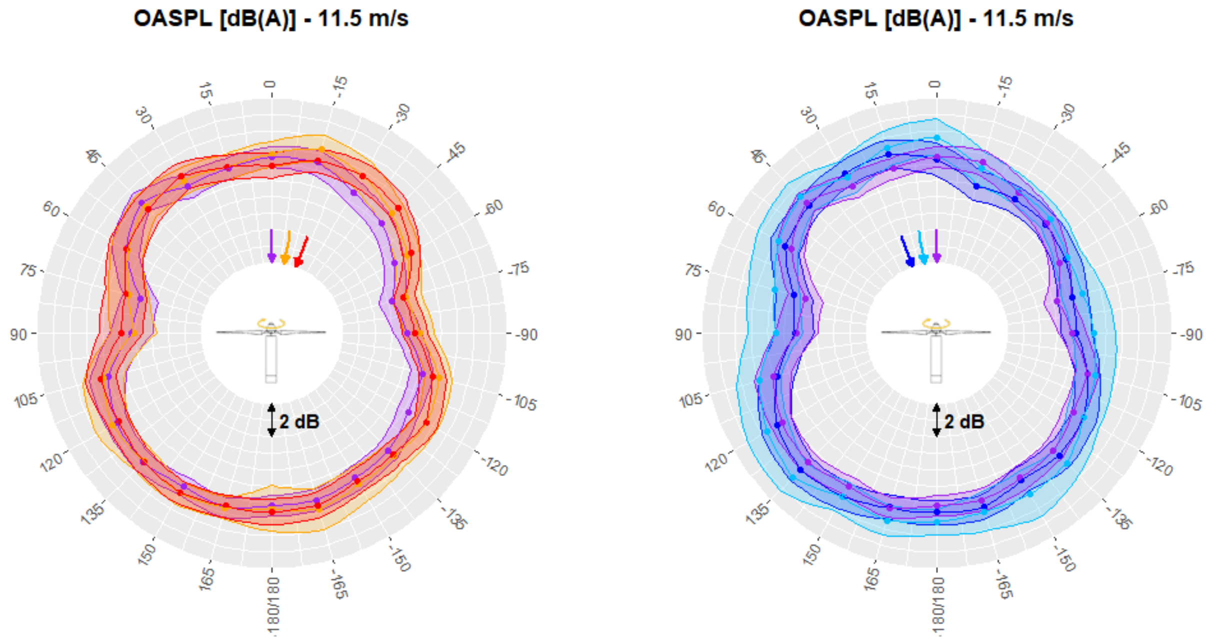

Figure 3: Sound pressure level (OASPL), displayed in the a fixed turbine reference frame. The colored arrows indicate the wind direction for each yaw misalignment case.

Figure 19: The current presentation obscures important trends. Plotting the change in noise emissions relative to the yaw 0-degree baseline as a function of wind speed would clarify how acoustic impacts vary across the operational envelope. The apparent result that moderate yaw offsets (±10 degrees) produce larger noise increases than extreme offsets (±20 degrees) at high wind speeds is counterintuitive and potentially significant. Does this non-monotonicity suggest optimal yaw angle strategies for minimizing acoustic impact while maintaining wake steering benefits? Provide mechanistic interpretation grounded in blade aerodynamics or turbulent inflow physics. ***We tried plotting the turbine sound power level relatively to the flow-aligned case (see Figure 4). We find this representation is not very clear, the curves and their uncertainty bands are much superimposed. Therefore, we think it is better to keep the Figure 19 of the manuscript as is.***

With this representation, it not very obvious that moderate yaw offsets (±10°) produce larger noise increases than extreme offsets (±20°) at larger wind speeds. Indeed, the -10° curve is a bit higher than the -20° curve, but its corresponding uncertainty band is also larger. The t-test run as an answer to question 4 above confirms that this difference is not statistically significant. Since there is not much difference between the +10° and +20° curves either, we prefer not to discuss these points.

1.3 Minor Edits and Technical Corrections

"Likewise" should be "Like" for grammatical correctness in this context.
Ensure consistent terminology for key concepts (e.g., "yaw misalignment" vs "yaw offset" vs "yaw control") throughout the manuscript.
Section 1, line 40: Remove extra comma in citation "(Göçmen, 2016), ."
Section 2.3, line 125: "Monday 25th" should specify "25 March 2024" for clarity, as the month is mentioned only in the preceding paragraph.

All these small corrections have been edited on the revised version of the manuscript.

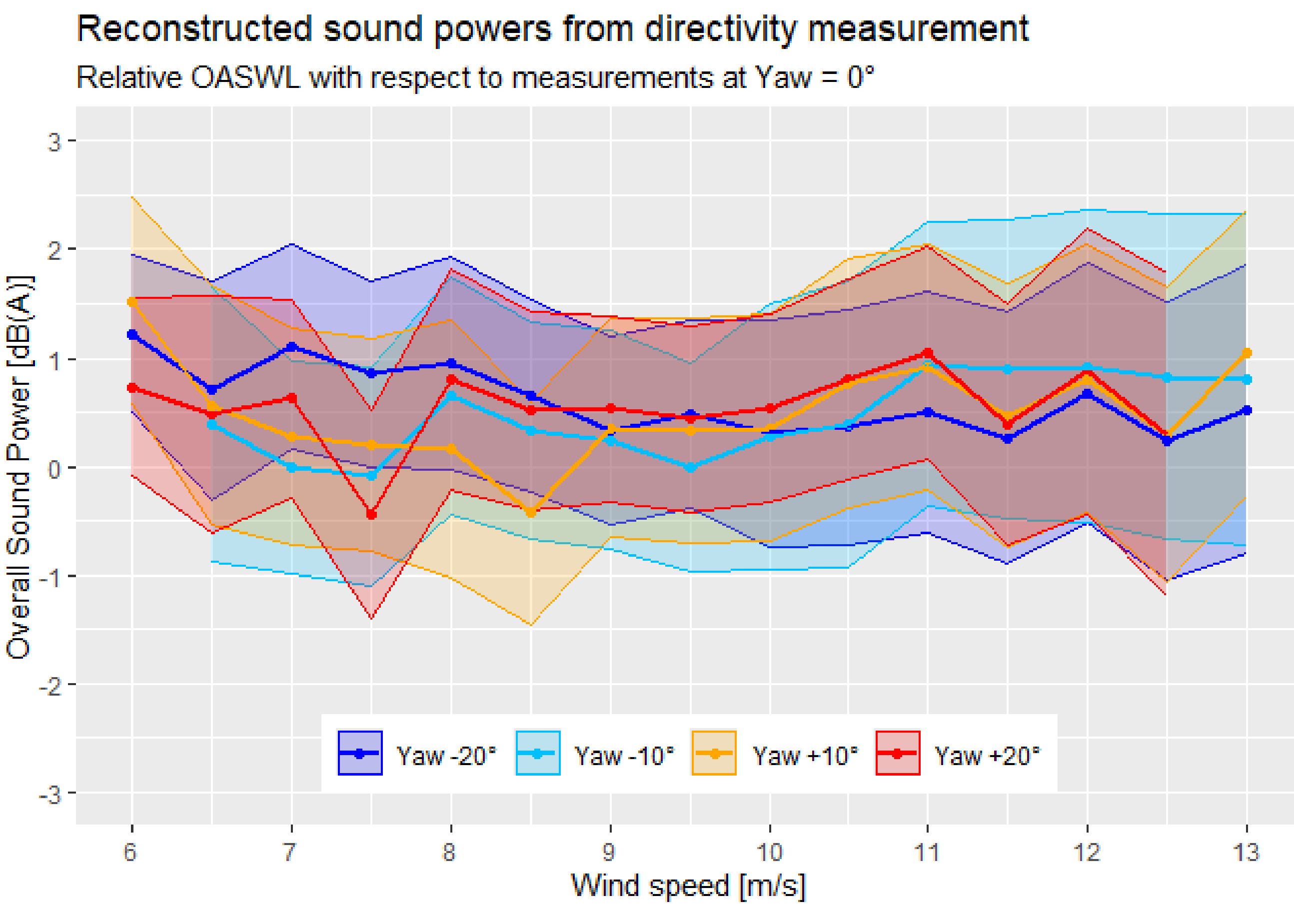

Figure 4: Relative turbine sound power level (OASWL) as a function of wind speed, with respect to measurements at Yaw = 0°.

2 Response to anonymous referee 2

In this study, the authors investigate the impact of wind turbine yaw misalignment on the noise emissions of a utility-scale onshore wind turbine. Using a circular array of 24 sound level meters (SLMs), a field campaign was conducted spanning four days that measured the acoustic impact of wind turbine operation in yaw-aligned and misaligned conditions. A nacelle-based and ground-based lidar were used to measure the wind speed, and wind speed measurements were synchronized to sound level measurements. The field measurements revealed asymmetries in noise emissions between the downstroke and upstroke sides of the wind turbine in yaw-aligned conditions, as well as between the downwind and upwind sides. In yaw-misaligned operation, a small but statistically significant increase of 0.6 dB(A) was observed, which contradicts previous studies that reported no change in noise emissions or a drop in noise emissions when a turbine is yaw misaligned. Overall, the experimental field campaign provides a useful and curated dataset to the flow control community. However, the postprocessing methodology includes several confusing decisions regarding time synchronization, wind and background sound measurements, and outlier detection. Furthermore, the main conclusions of the article contradict previous field campaigns on the same subject, with minimal discussion on possible sources of discrepancies leading to the contradictory results. Specific comments are outlined below.

Thank you very much for you thorough review of our manuscript. As you can see in our comments below, we tried to clarify our methodology with respect to the data processing and improve the discussion regarding our results and the comparison with the state of the art, hopefully improving the general quality of the paper.

2.1 Major comments

1. The primary result the authors claim is that the overall A-weighted sound power level (OASWL) of the yaw misaligned turbine is higher than its yaw-aligned counterpart, flow conditions remaining the same. However, this result contradicts that of previous studies, and the discussion regarding dicrepancies is very limited (one sentence at L449). The authors should thoroughly explore differences in their campaign that could lead to a contradictory result (distance to rotor, size of turbine, wind speed ranges, amount of data, pre/postprocessing strategy, etc.).
A section has been added at the end of the paper to discuss results reported in the literature and to propose some interpretation for the apparently diverging conclusions. In addition to this section, rotor speed curves in Figure 1 show that the curve in misaligned conditions are generally not below the $\beta = 0^\circ$ ***case (except occasionnally Yaw +20°). This is a possible reason for discrepancies with the main reference of NREL.***

2. Throughout several steps of the data postprocessing, it is stated that data are "visually inspected" (e.g., L115). This is also mentioned in the outlier detection (see next comment). Additional details should be provided regarding the data processing methods.
Both phrases were reformulated to provide additional details on SCADA signals considered for the visual inspection, and refer to Figure 8 for the case of a noisy event.

3. The outlier detection uses a threshold value of 10 dB to filter out observations relative to a ±50-second running median. However, figure 10 showcases the deficiency here: one spike in SPL (around 21:14) is removed as an outlier, but a very similar looking spike around 21:42 is not removed. What is the sensitivity of the results to the outlier removal strategy? How do the results change if other values for the running median window and/or dB threshold are modified? And how were the values in the study selected?
We added in the paper a justification of the selection procedure for the outlier detection in section 3.1.2. A sensitivity analysis to outlier detection has been conducted using a much more sensitive filter with a threshold value of 5 dB to filter out observations relative to a ±25-second running median, with a safety window of ±20 s. The total number of outlier events raised from 2150 to 4616. Sound level changes are very limited as can be seen on the directivity patterns on Figure 5. Minimal variations are observed as well concerning the estimated sound power levels on Figure 6. The mean absolute variation stays below 0.04 dB between the two filters. Other parameter values have been

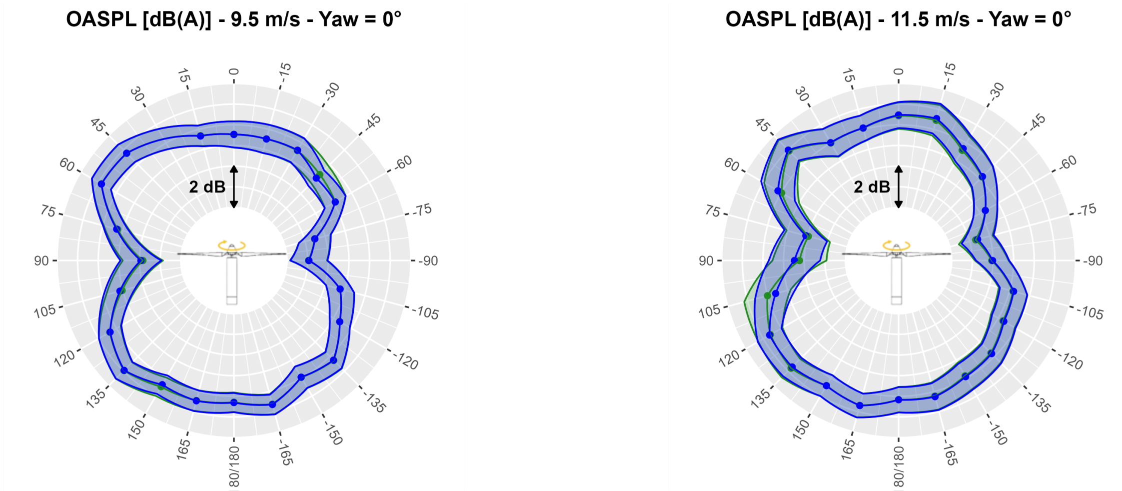

Figure 5: Sensitivity of directivity patterns to outlier filter. The filter used in the paper (with a level threshold at 10 dB, from a ±50 second running median, with a safety window of ±20 s) is compared to a more restrictive filter with a level threshold at 5 dB from a ±25 second running median with a safety window of ±20 s.

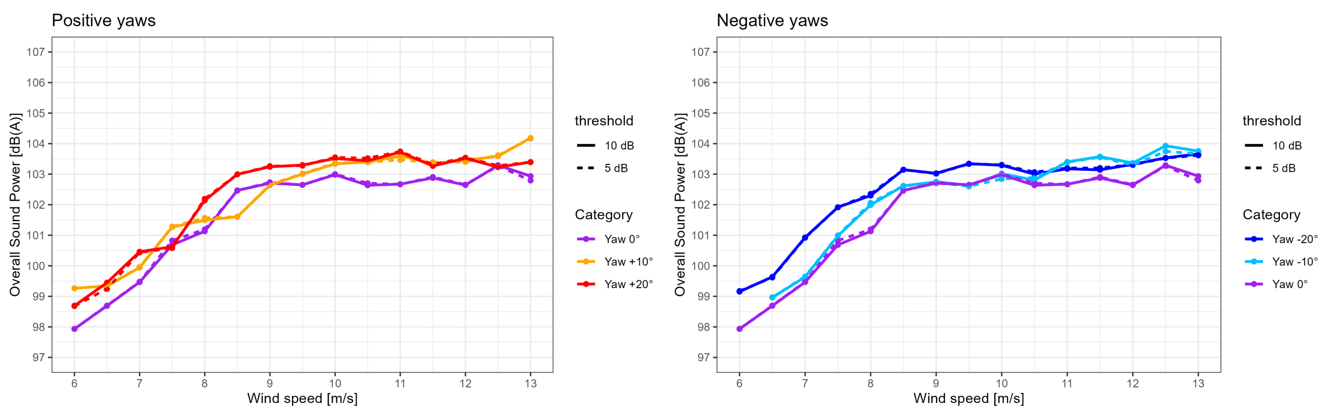

Figure 6: Sensitivity of estimated sound power levels to outlier filter. The parameters are identical to Fig.5

tested provided the same order of magnitude.

4. The study would benefit from a specific discussion regarding the measurement of the freestream wind speed. For example, what is meant by "the resulting wind speed signal [from the ground-based lidar] was too noisy and did not correlate well with the SCADA data"? Typically, SCADA data are output in 10-minute frequencies, however the study here seems to aim at 10-second time intervals. Do the various methods agree well in yaw-aligned flow, or when the background noise measurements are active? If not, then that would raise concerns for using different measurement channels throughout different experiments (yaw-misaligned vs yaw-aligned vs off).
Truly, SCADA data are typically output in 10-minute resolution, however the application if the IEC 61400-11 standard requires the recording of SCADA data at 1-second frequency and aggregation of these data into 10-second chunks. We therefore had to process lidar data into 10-second chunks as well in order to be able to associate them with the SCADA and acoustics data. Several methods were developed for that objective, and compared to the active power reference signal in yaw-aligned flow. We cannot use the background measurement periods for this kind of comparison : indeed, since the active power is null the power curve cannot be inverted, and as pointed out in the IEC noise standard the nacelle anemometer measurement is likely to be biased when the turbine is stopped. The Table 2 in the manuscript now recaps the correlation between the different available lidar wind

speed signals: it can be seen that the one chosen for the study is the best performing by far. A new paragraph has been added to discuss those results.

5. Section 3.3.1: Is wide variation in SLM background noise physical, or could be due to calibration error in the SLM? If it is physical, then this suggests there are much larger variations (upwards of 4 dB) in small variations in location around the wind turbine (¡100 m), compared with relatively small differences (order 0.5 dB(A)) in noise due to asymmetries and yaw misalignment.
The SLMs were calibrated every day before the measurements begin and did not show any inconsistencies during the 4 days of measurement (except for the case of SLM #06 that is mentioned in Section 3.1.3). We therefore think the wide variation in SLM background noise is rather physical. There is indeed much larger variations in due to the location of the SLM compared to the effect of yaw misalignment, however since we calculated in the methodology a specific background noise for each location, we expect that these variations cancels out when we derive the turbine specific noise. We added some text in the manuscript to discuss a bit more these variations of background noise and justify our choice for a specific background noise calculation by SLM.

2.2 Points comments

1. Figure 1: As I understand it, this decision matrix describes acoustic curtailment for a specific four-turbine wind farm. How general is this decision matrix? Additionally, the times of curtailment center in the evening and nighttime hours, while this experiment is performed only during daytime hours. How might that affect the results or conclusions?
The lookup table presented in Figure 1 is not the one that is implemented at the farm. It displays a very typical case of acoustics curtailment in France, since they are generally defined according to Daytime/Nighttime, wind sector, wind speed and season of year. This is done to account for specifities of the site both in terms of climatology (e.g. long-term wind rose), acoustics propagation, local regulatory constraints and background noise observed at the location of people habitation (i.e. a few hundred of meters away from the turbine). The measurements that we describe in this manuscript follows the procedure described in the standard IEC 61400-11 and are realized in the close vicinity of the turbine in order to characterize its specific noise only, independently of all those aspects. Therefore, the fact that our measurements were realized during the day, and curtailment in France generally applied during the night does not change our conclusions. Of course, any knowledge gained thanks to this measurement could in turn be used to update some inputs of the modeling toolchain that is used to optimize our acoustics curtailment, but this will be the subject of future work (some of it will be realized in the scope of TWAIN project).

2. L94: the terrain inclination varies between 29° and 31.5°. Perhaps I am misunderstanding how inclination angle is computed, but this does not seem "very simple and flat". Clarity would be appreciated.
This sentence refer to the terrain inclination angle ϕ defined by the IEC 61400-11 standard (see the figure 4 of the standard) as :

$$\phi = arctan\left(\frac{HH + dz}{HH + D/2}\right) \tag{1}$$

where HH is the turbine hub-height, D the turbine diameter, and dz the difference in altitude between the turbine and the SLM location. For a perfectly flat terrain ($dz = 0$ m) and the corresponding turbine dimensions ($HH = 80$ m and $D = 110$ m, this would yield $\phi = 30.65°$). Therefore with ϕ varying between 29 and 31.5° depending on the SLM location, we are close to be in the ideal case of the flat terrain. The phrase in the manuscript has been reformulated to clarify how ϕ is defined.

3. L126: It may be worth noting that high windspeed events (i.e., above-rated conditions) and high turbulence conditions are less important to WFFC applications. What were the wind speeds during this period of no testing?
This is a good point, we added a sentence in the manuscript to point that out. The wind speeds when

we were forced to stop the 6^th round of measurement were above 15 m/s.

4. Figure 4: How is vertical shear computed?

The wind shear is estimated by fitting a power law to 14 planes of measurement ranging from 40 to 140 m (i.e. the span of the turbine rotor). This clarification has been added in the manuscript.

5. Figures 6-7 may be combined into one figure to save space. Furthermore, the names of the SLMs should be changed in figure 7 and throughout (e.g., figure 8) to more useful names (such as "SLM #XX" as used throughout the text).

Figures 6 and 7 are one-column figures, in the discussion manuscript that leaves a lot of blank space but in the final version (two column display) they will end up side by side. All figures have been reedited to harmonize the SLMs names in all of them.

6. Regarding the clock drift correction, what is the point of steps 1-3 (section 3.1.1), when step 4 seems to align all of the time signals with the turbine operation/SCADA clock?

By considering first all microphones for estimating the clock drift correction, we calculate a time offset which is based on multiple pairs of microphones and that can therefore be considered more robust and if it were just obtained though a direct comparison to a single SCADA signal. A sentence has been added in the manuscript to clarify this point.

7. L212: Was the dB offset correction to SLM #06 calculated during turbine operation, or only during background noise measurement periods? If the offset correction was computed for the entire time history, then it seems like that could bias the results.

The offset correction was calculated for this entire time history. This could indeed bias the result. We decided to discard the data of SLM #06 for the first measurement day and updated all the figures in the paper. The impact of this change is negligible and does not change at all the mains conclusions of this research.

8. L251-254: Please clarify the method for measuring wind direction at hub height. What "external sensors" are referred to in L253? What averaging or interpolation methods (if any) are used to measure the hub height wind direction with the ground based lidar?

The two external sensors considered here are the nacelle-mounted and the ground-based lidars. The 10-minute wind direction signal, averaged over all planes of measurement of the ground-based lidar is taken as reference for the wind direction. This precision has been added to the manuscript.

9. Table 3: The yaw angle uncertainty/error is missing from the sources of error. Previous field campaigns (e.g., McKay et al., Wind Energy (2013)) have noted significant discrepancies in turbine alignment from the incident wind, due to slow-reacting controllers and turbulence.

This is a good remark. There is indeed an additional source of uncertainty due to the wind dynamics and slow reaction of the turbine yaw controller to changing wind direction. However, our dataprocessing is based on the recommendations of the IEC 61400-11 standard, which does not consider this source of uncertainty and since we did not have the means to estimate it accurately we decided to keep the same categories as the ones listed in the standard. A comment and a reference to this citation have been added in the manuscript with respect to this point.

10. For this audience, it would be useful to have a brief discussion clarifying between OASPL (pressure levels) and OASWL (power levels). Why is OASPL shown in figures 14-18 but then OASWL shown in figure 19?

An explanation of the difference between SPL and SWL has been added in section 3.3.3, together with a clarification of the term* overall*. OASPL is shown in figures 14-18 because they show local values (noise levels measured at each microphone positions), while figure 19 show integrated values over all microphone positions. OASWL estimates the global acoustic energy flux radiated by the turbine in all directions at once.

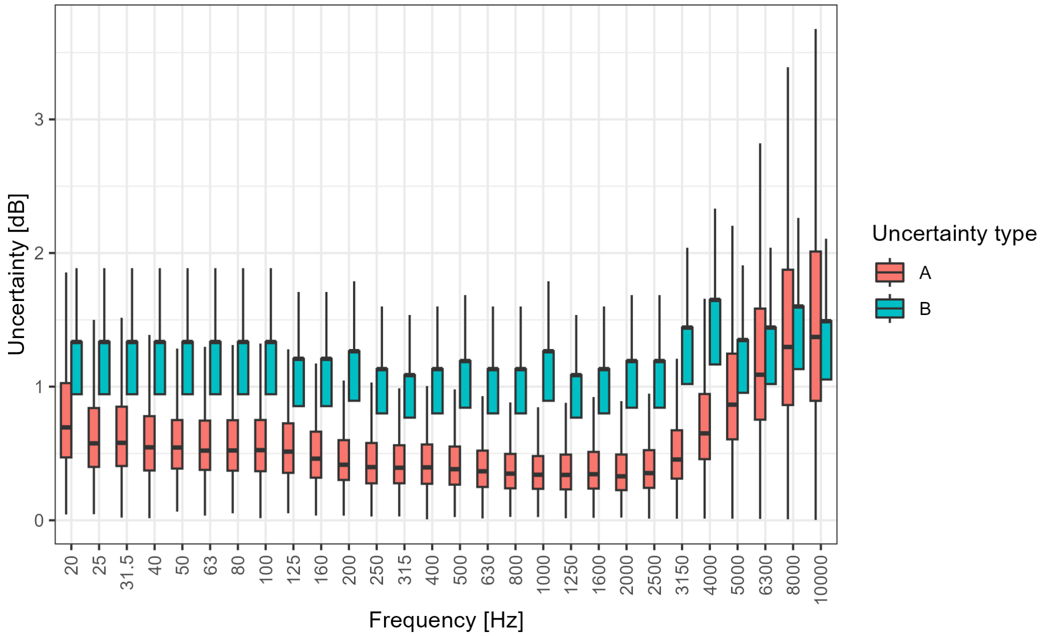

Figure 7: Uncertainty type statistics for total noise as a function of 1/3-octave frequency bands. Type A is related to the statistical dispersion of measured data while type B is related to estimated uncertainties of material and instrumentation.

11. L383-384: How does the discussion of airfoil camber/asymmetry relate to the data collected in this experiment and/or this wind turbine model?
The discussion on camber has been completed to finalize the reasoning on directivity pattern assymetry. We do not have the detail of the blade geometry so the discussion stays on a general level, without quantitative assessment from this specifc experiment.

12. L394-395: where does the increased uncertainty at higher frequencies come from? As I understand it, the uncertainties in table 3 are all constant with frequency/octave band.
Concerning table 3, u_{B2} and u_{B4} are frequency dependant so "Type B" uncertainty at 5 kHz is indeed higher than at 500 kHz. This can be observed in Figure 7. However the main variation in uncertainties across the frequency range is provided by "Type A" contributions, related to the natural variation of sound levels. A sentence has been added in section 4.1.

13. Figure 19 and surrounding text: the uncertainty bars overlap significantly, yet the p-value is small. Why is this?
This is because the p-value is an aggregated indicator of the whole curve, over all windspeed bins. Uncertainty bars only refer to a specific wind speed bin. A p-value < 0.05 indicates that it is highly unlikely that all these points at $\beta = 0°$ lie below the other curves due solely to uncertainty (as quantified by the error bars).

2.3 Minutiae/typos

1. L24: A citation to Meyers et al. Wind Energy Sci. (2022) would be particularly relevant here.
Very good point, we added the citation.

2. L51 and throughout: what does TWAIN stand for? Is it an acronym?
TWAIN is the short name for Integrated, Value-based, and Multi-objective Wind Farm Control pow-

ered by Artificial Intelligence. Since the project name is quite long, and the acronym for TWAIN not obvious, we did not detail in the manuscript the full project name but instead just added a link to the project website in the references.

3. L89, L104, and throughout: some writing is too informal for an academic publication.

These sentences and a few other in the manuscript have been reformulated.

4. L144: Please be more specific than "Unfortunately some sensor defaults introduced a loss on data in some conditions."; this sentence (and others like it in the manuscript) should be omitted.

This sentence have been rephrased.

5. Figure 5 caption and elsewhere (L155, L276): "10-seconds sample" → "10-second sample".

These typos have been corrected.

6. The opening paragraph in section 3 is fully described in figure 12; consider moving figure 12 earlier and rephrasing the intro text in section 3.

Thanks for the suggestion. Indeed, it makes more sense to move the figure in the intro of section since this also how it is done in the standard. The text have been rephrased accordingly.

7. L301: to clarify, is the OASPL also binned on wind speed, in addition to relative position and yaw angle?

The mention to wind speed has been added.

8. Figure 16, y-axis: "OAPSL" → "OASPL"

The typo has been corrected.

9. L472: The acronym SWL is never used in this manuscript.

This unused acronym as been removed.

10. L476: "first of his kind" → "first of its kind"

11. L476: "WFC" → "WFFC"

Both typos have been corrected.