Review of manuscript: WES-2021-153
Title: Comparison of Large Eddy Simulations against measurements from the Lillgrund offshore wind farm
Authors: Sood et al.

The submitted paper focuses on a detailed comparison between large eddy simulations (LES) of the Lillgrund offshore wind farm and experimental LiDAR and SCADA data. Overall, I am impressed with the technical challenge the authors have approached. Further detailed comparisons between LES and real wind farm measurements are needed in the literature to further establish confidence in farm models, and to identify areas which require modeling improvements. The authors have provided the appropriate detail regarding their advanced LES solver and wind turbine representation, which includes turbine rotation and loads. Finally, the field data itself is comprehensive and compelling.

However, I do have several comments and suggestions which I encourage the authors to consider in a revision. Principally, while significant attention is placed on the setup of the study (~20 pages including Appendices), there is less attention to detail in the analysis of the results, the uncertainties of the analysis, conclusions, and future work (~8 pages which are mostly figures, see point comments below).

I suggest that the authors add a new section which outlines the uncertainties, gaps in knowledge/modeling, and highlights the future work that this study motivates. There appear to be many areas which need to be improved before sufficient accuracy is obtained in LES wind farm modeling, especially in ABL representation. Several figures can also be improved in their visibility.

In the abstract, the authors state that ‘good’ agreement is obtained between LES and field measurements for power, loading, and wake recovery. I suggest to include quantitative descriptions of the error and to eliminate the qualitative descriptors, since different readers may have different take-aways from the results (e.g. looking at Figure 11, there appear to be statistically significant differences in the power between LES and measurements for around half of the turbines in the farm, which I would argue is not a good agreement). I also appreciate that the authors refer to some errors in the abstract, and the need for improved controls modeling and a finer grid.

Finally, especially given the preliminary nature of the comparison between model results and field LiDAR data, I encourage the authors to consider publishing the wind farm LiDAR and SCADA measurement dataset along with their paper. This would be valuable to the community.

<u>Point comments:</u>

1. Abstract, Line 10: What is the optimization framework mentioned here? To this point, optimization hasn’t been mentioned in the abstract.
2. Abstract: Quantitative results in the abstract would be useful to add.
3. Line 46: Typographical error.
4. Figure 1:

   a. Overall, I found this figure to be a bit confusing. Consider labeling elements within the figure, in addition to a discussion in the caption. More information could also be added to the caption (e.g. defining the red and blue arcs, which are not currently defined).
   b. Define Vara, Levante, Sternn.
   c. The turbine number labels for the turbines which have a mounted LiDAR are missing.
   d. Consider adding an upwind facing arrow to the magenta inflow centerline. Please also define it in the caption.
   e. Why are the red and blue scan areas chosen not symmetric?
   f. More details are required to explain the "transect lines". The Figure mentions plural "transect lines" but I see only one line?
5. Line 78: Was the turbine nacelle direction only of 'Vara' used to characterize the true inflow direction or was averaging over multiple turbines performed?
6. Line 81: How high above sea level was the LiDAR mounted?
7. Section 2.2.2 was difficult to follow. What are the "three transects"? Is the third transect the upwind inflow LiDAR transect? Why was 10-minute averaging selected? Is this synchronized with the SCADA data?
8. Line 96: Did the processing to result in uniform 0.5 Hz sampling involve upsampling or downsampling?
9. Line 99: Why was the nacelle calibration performed using wake losses (assuming that the wake advects exactly with the mean wind) for the SCADA data but it was performed using nacelle direction measurements for the inflow LiDAR? What magnitude of uncertainty do the authors estimate is incorporated through these correction methods?
10. Line 99: Typographical error.
11. Line 126: *"The effect of the sea surface is included using a wall-stress model [...]"*
    At first, I found this to be confusing because as written it seems to imply that the effect of waves are considered (but that is just my interpretation), but this is a standard wall model. Consider rephrasing the sentence?
12. Sections 3.2 and 3.3: These sections (along with Appendices) are clear and detailed. Have the authors previously compared this AASM model to ALM, ADM, etc? There are a growing number of actuator turbine representations (e.g. actuator surface). It would be helpful to add a sentence or two to concisely summarize pros and cons of the selected AASM methodology.
13. Figure 4: I think this figure is really great, very clear. Why was a first-order time-filter selected? Has this decision been validated in previous papers/theses by the authors for the AASM? If so, please provide references and justification. Also, consider defining variables in the figure caption.
14. Figure 5: Are the deviations in this figure the result of a difference in the induction given the coarse grid resolution in SP-Wind?
15. Figures 6 and 7: I presume that the individual data points shown are not the full resolution of the precursor LES cases (e.g. CNk8 has two points in the boundary layer). Please clarify.
16. Section 4.1:
    a. Does u^t represent the velocity vector (it is not bolded and no vector identification is included) or wind speed? If it is wind speed, is the spanwise velocity (v) neglected from

this analysis? If it is a vector, are there unique values of $u^{*,t}$ and $z_0^{\,t}$ for each of the streamwise and spanwise velocities?

   b. How do the authors replicate the wind veer from the LiDAR? This is not included in Eqs. (5) or (6).
   c. Also, e1 is not defined in Eq. (5).
   d. This discussion and the references focus on surface layer similarity. Cases with geostrophic forcing at a finite Rossby number are also considered. Matching the friction velocity can lead to large errors in the geostrophic wind speed (as seen in Figure 8).

17. Line 225: What does it mean for the range gate to span the hub height, as the hub height is one location.
18. Table 3: What is the uncertainty associated with the TI estimation for both LiDAR and LES?
19. Figure 8 is messy with the inset figures overlapping the outer figures. Please consider reformatting. Also, how were the 95% CIs estimated for both LiDAR and LES? Please add this description to the text body.
20. Section 4.2: Overall, I find this section and the outlined approach interesting. I recognize the technical challenges of what the authors are attempting, given fundamental inconsistencies between the CN or PD LES cases and the true ABL LiDAR measurements. Primarily, fitting neutral or PD ABL profiles to non-neutral ABLs may bias the results. So, I suggest the authors add an additional paragraph, addressing the following:
    a. Discussion of the stability during the LiDAR measurement time periods.
    b. Discussion of how the optimization process which finds the minimum error compared to conventionally neutral or pressure driven boundary layers (with no stratification) may bias the considered cases.
    c. Given the poor agreement in the wind veer, discuss how the optimization process may have biased the results to over-emphasize wind speed compared to wind direction.
    d. Discussion of how the pronounced disagreement between LiDAR and LES above the rotor area may bias the vertical transport into the wind turbine array atmospheric boundary layer (WTABL).
    e. Finally, I find it a bit strange that several of the selected cases had LiDAR or loads measurements missing. Why did the authors not select other cases where all of the data were present (even if the precursor data differed slightly more than in other cases)?
21. Line 286: *"This is in accordance with the errors in the inflow profiles observed [...]"*
    The analysis of the power results in this sentence appears weak and selective. First, putting aside the power and looking just at the wind profiles in Figure 8, all of the PD cases have high error in the wind veer, not just PDk3. Second, it is not clear from Figure 8 that the PD cases have lower wind speed error than the CN cases. For example, CNk8 seems to have fairly low wind speed and wind veer error. Cases PDk1 and PDk2 do not appear to have significantly lower wind speed error compared to the LiDAR than CNk8. The authors should revisit this analysis with more attention. I suggest quantitative metrics to support the statements.
22. Figure 10: How sensitive are the bootstrapped CIs to the block length selected in the bootstrapping approach?
23. Figure 11 is challenging to see. Please modify the aspect ratio to be wider and shorter.

24. Line 304: Do the LES turbines align themselves with the wind direction measured locally at each turbine?
25. Figure 13: I believe the y-axis of this plot should be 'Turbine orientation' or 'Nacelle position' rather than wind direction. In the field, the turbine's orientation is not always aligned with the wind direction, depending on the yaw controller dynamics. The local wind direction may also vary within the farm. Also, why are the turbine position errors high for all turbines in PD1? Does this indicate that the calculated mean wind direction was not correct?
26. Line 310: *"This indicates [...]"*
    This sentence seems to contradict the previous sentence.
27. Figure 14: The difference between measurements and LES for the star symbols is not clear in these figures. The open star looks almost filled.
28. Line 323: *"For the PDk2 case [...]"*
    Again, I encourage the authors to be more precise in their analysis. For PDk2, with the errorbars in Figure 14 representing 95% CIs, there are statistically significant errors between LES and LiDAR for B08 and B06, while the differences between LES and LIDAR for B07 are not statistically significant. This differs from the present analysis discussion in the paper.
29. Figures 16 and 17: Why is the far wake comparison omitted from these plots? Please include. They are mentioned on line 335, so presumably the data is there. If they were selectively omitted due to poor agreement between LES and LiDAR, that is even more reason to include them in the figures.
30. Line 337: *"Observing the wakes [...]"*
    I was not able to discern the turbine orientation story from the LiDAR results in figure 17(a). Could the authors rephrase this discussion to be more clear? Looking at figures 17(a) and 17(b), how is it clear that in LES there is partial wake while in LiDAR there is not, since only C05 can be seen (and it does appear to be in a partial wake).