Author response – Minor revisions

Overall, I am impressed by the ambition of this paper which combines many years of incremental progress in improving LES simulations of wind farms into a single comparison between LES and field data. My main takeaway from the predictive accuracy of the LES is that progress has been made, but there is a lot of room for improvement (e.g. meso-microscale coupling rather than simplified boundary layers which approximate the wind speed profiles), as there are significant differences between LES and field data in both power and loads for most turbines within the farm. I hope in their final revisions, the authors can help contextualize how their results can be useful to guide future research.

We thank the reviewer for taking the time to review the manuscript, and acknowledging the importance of the performed study. Through the detailed analysis performed in this work, we aim to highlight the strengths and weakness of LES solvers in predicting performance of offshore wind farms. We have taken the reviewer's valuable feedback into consideration, and have made the following changes as per the point comments.

Point comments:

I appreciate that the authors have added much more quantitative analysis and discussion in their revised paper, I think that adds clarity to the results. Yet, I believe the following statement in the abstract is still not an accurate summary of the results presented in this paper.

"Nevertheless, statistically significant errors were observed on an individual level for a few turbines deeper in the wind farm operating in wakes across the simulated cases."

As per the reviewer's suggestion below, we have performed an analysis to quantify how many turbines for each case have statistically significant errors. We observed that this proportion was higher for the cases which have a larger number of aligned turbines, and in general it was the waked turbines that exhibited statistically significant errors. The above line the abstract is thus modified to be "**However, when comparing individual turbine power production, statistical significant errors were observed for 16% to 84% of the turbines across the simulated cases, with larger errors being associated with wind directions resulting in configurations with aligned turbines**"

Taking a look a Figure 12 (I have counted by hand so I may be off by a couple), 32 of the total 48 turbines (32/48, or 67% of the total turbines) have statistically significant differences (error bars which don't overlap) between the LES power predictions and the field data. I encourage the authors to perform this quantitative analysis more formally than I just have, and to add these quantitative results to the abstract.

We thank the reviewer for this suggestion and have performed this quantitative analysis for all the cases. The results are summarized in the table below and have been included in the manuscript in section 5.3.

Case	Turbines with significant errors [%]
PDk_1	58.3
PDk_2	16.6
PDk_3	83.3
$CNk4_1$	25.0
$CNk8_1$	35.4

I want to again emphasize that I believe this paper and the methods are of high quality and are a useful contribution to the literature. But I don't think we can overstate the fidelity of agreement between high quality microscale LES and field data (recommendations also echoed by the other reviewer). There appears to be a long way to go before we have reliable LES predictions of wind farm power (much research to be done!). This paper can be an excellent summary of key technical gaps.

We thank the reviewer for their recognition of the research and methods presented in the manuscript. We have included the following lines in the conclusion section to highlight the technical gaps highlighted in the paper. "**The analysis in this work thus highlights certain areas in wind farm LES which still require further research to make accurate performance predictions for offshore wind farms. In particular, lack of prior knowledge of meso-scale effects, field turbine control logic, and insufficiently fine grid resolution are highlighted as significant actors which can lead to errors in predictions. Therefore, future work should focus on improving on these effects , in particular for situations with aligned wind turbines**"

Section 4.3 is well-written and useful!

We thank the reviewer for their kind words.

Line 373: As in the first comment, this sentence is worth rephrasing from "good agreement with field data" to "reasonable qualitative agreement with field data, [...], but statistically significant differences for 32/48 of the turbines considered."

As per the performed analysis, this line has now been modified to "**However when comparing individual power production on a turbine level, statistically significant errors are evident for 58% of the turbines for the PDk1 where error bars for the 95% confidence intervals do not overlap when comparing LES predictions and field measurements. The same analysis is performed for all the simulated cases, and the proportion of turbines with statistically significant errors are presented in Table 5. In general, it was observed that the turbines at the front of the farm did not have statistically significant errors, however turbines operating within the wakes of an upstream turbine had higher errors which were significant for wind directions resulting in an aligned configuration**."