Review of wes-2025-286 titled "Impact of Boundary Layer Height and Large-Scale Turbulence on the Efficiency and Loads of Offshore Wind Farms" by S. Ivanell *et al*.

General comment: The authors examine 5 offshore sites in Northern Europe to establish a connection between boundary layer height (BLH) and wind farm efficiency. The study, conducted using both WRF mesoscale simulations and Lidar observations (for one of the sites), reveals higher wind farm efficiency for strong wind and high BLH, as opposed to lower BLH and/or wind speed. Additionally, turbulent large scales typically present in offshore flows are shown to have an impact on fatigue loads beyond the temporal duration typically considered by fatigue design standards. The topic is of primary interest both for WES and for the wind energy community in general, given the ongoing growth of offshore wind turbines, both in size and number.

Before recommending this article for publications, a number of major points should be addressed. On a high level, I had the following impressions reading the manuscript:

- Some details regarding the technical approach are overseen and left to the reader's intuition (see, e.g., lines 235-239).
- The authors are over relying on previous publications when it comes to provide insights on the technical approach. The details about what is done in this article should be clear just by reading this article. More interested readers can find more information by looking at previous publications, but numerical simulations and experimental setups should be described here in more detail.
- The technical details are provided in a quite scattered fashion. Some details regarding Lidar scanning strategies are left to the Results section (lines 250-254, for example), while some results are described in the Methodology section. I suggest devoting one sub-section to numerical simulation details and another one to experiments.

More comments are reported in the remainder of this document.

Specific comments

Lines 62 – 64: The approach proposed by the present study is a bit confusing at this stage of the manuscript. Please clarify further how you map different atmospheric conditions and how you connect them to climatology.

Lines 102 – 104: The presence of large-scale turbulent motions is already a well-documented feature of offshore environments (see, e.g., ref.), so reiterating this point in the present study is not necessary unless it is better specified for the present scope.

Lines 114 – 117: These lines could be nicely integrated with a figure reporting the position and scan direction of each Lidar.

Line 135: Which parameter and value are the authors utilizing to define neutral atmospheric conditions?

Line 149: The necessity of utilizing both 2D and 3D spectral models is still unclear. Also, it appears that ψ and Γ exhibit significantly different behaviors, i.e. ψ decreases with velocity and shows little sensitivity to thermal stability, while Γ increases with velocity and shows distinct values according to different stability classes. Please clarify.

Lines 159 – 160: Similar to the discussion about Mann's spectral model, please report the explicit formulation of 2D and 3D spectral models, in addition to a brief description of their parameters. Also, at this stage the rationale of using two different spectral models for the same dataset is unclear.

Line 164: Since the authors mention climatology assessment, I expect the measurements to cover a multi-year period. For FINO1, this period corresponds to 2 years (if I understood correctly), which is reasonable. Please report the analogous measurement periods related to the remaining sites.

Section 3.1: In my impression, this section is a repetition of what is already displayed in Fig. 5 and 6 without any real insights of physical phenomena explaining the observed differences at sites A, B and C. Please either provide more insights or remove this section.

Line 204: Is the time local or UTC zone? Also, which method(s) did the authors use to evaluate BLH from Lidar data?

Line 205 – 206: The peak of occurrence in Fig. 6a-6c seems to occur during daytime convective conditions, since it's characterized by low wind speed and high BLH. There should also be a secondary peak related to stable night-time conditions featuring low BLH and high wind speed. Please provide more context for this.

Line 219: From Fig. 7, it appears that the overall wind farm efficiency is very close to 1 (and even greater than 1 from Tab. 2, 3, 4 and 5), thus meaning nearly zero under-production from the second to the last row (cf. Eq. 1) and zero blockage. This seems a contradiction with existing literature. Please provide further clarification.

Figure 7 and Tables 2 – 5: Please revise the unit on the x-axis ($m.s^{-1}$).

Line 233: I am a bit skeptical about the extrapolation of the efficiency outside of the simulated bounds, especially since it yields values larger than one for high BLH and wind

speed. Perhaps you could utilize a subset of the simulated points and extrapolate on the remaining ones to assess the accuracy of this method.

Lines 237 – 238: There are a few caveats with this statement. First, please provide more details on the calculations of power from C_p and thrust (did you use C_p and C_t curves?); second, the length of the climatology period (5 years) has not been stated previously, and it should be reported where the simulation setup is described. Third, possible discrepancies between thrust and C_p -based estimates of power should be discussed.

Lines 239 – 240: Please revise this sentence as it does not sound grammatically correct.

Line 240: A1 - A4 are not appendices, they are figures listed in Appendix A without neither context nor explanatory discussion. This falls below the standard of acceptable scientific publications. Please either add text to Appendix A or remove it.

Lines 242 – 243: This sentence is hard to understand. How do the authors quantify stable conditions? Also, the farm blockage typically refers to the slowdown exerted by the entire wind farm on the flow upstream, whereas here it seems like the authors are using the term "blockage" to refer to wake interactions or single-turbine blockage.

Line 244: According to previous statements, the total efficiency is the product of local and non-local effects, not their sum. Please revise.

Table 6: This table is not mentioned in the text; hence I suggest either remove it or add it to the discussion. Also, it makes little physical sense to report efficiencies with 6 decimal digits, unless the reported values are equal up to the 5-th decimal digit.

Lines 250-254: Again, this information pertains to the description of the experimental campaign, not the Results section. Additionally, identifying the mixed layer height with the BLH is valid only for daytime well-mixed atmospheric conditions. How did the authors treat nighttime stable conditions?

Line 256: Table 7 reports data from February through June, so it does not fully cover winter and summer.

Table 7: Is the Average of Bias a percentage or is it reported in meters? Please clarify.

Line 269: This is a crucial passage of the uncertainty quantification that deserves more in-depth analysis. The authors utilize the values plotted in Figure 11 as a transfer function ($f(BLH, U)$, where U is the horizontal wind speed) to map the BLH bias (WRF vs. lidar) into wind farm efficiency uncertainty, thus relying on the equation:

$$\eta \approx f(\mathrm{BLH}, U) * \mathrm{BLH}$$

However, I also expect a bias between the WRF and Lidar-derived values of wind speed. Assuming Gaussian distribution of BLH and wind speed bias, the previous equation becomes:

$$\Delta\eta \approx \frac{\partial f}{\partial \mathrm{BLH}} * \Delta\mathrm{BLH} + \frac{\partial f}{\partial U} * \Delta U,$$

where quantities indicated by Δ indicates the interquartile difference. If the wind speed bias is negligible, please report literature references or, otherwise, please quantify it.

Line 275: The largest uncertainty of the wind farm efficiency is found for relatively low BLH values, which correspond to night-time conditions. I would be cautious in considering Lidar-derived BLH as ground truth for stable, night-time conditions. A significant portion of the observed BLH bias could be introduced by the instrument and, thus, not related to WRF modeling uncertainties. I wonder how the Lidar-derived BLH uncertainty during nighttime compare with the total uncertainty.