This study investigates the impact of atmospheric boundary layer height (BLH) on wind farm efficiency and the influence of large-scale turbulence structures on turbine loads. The authors combine mesoscale modeling (WRF) with lidar measurements to assess climatology and uncertainty across sites in the North Sea and Baltic Sea. They further utilize CFD (code_saturne) for efficiency estimation and aeroelastic simulations (HAWC2/DYNAMIKS) for load analysis. The findings suggest that lower BLH reduces efficiency and that large-scale turbulence increases fatigue loads due to wake meandering.

Major Comments

1. The manuscript in its current form lacks a strong narrative thread. It currently presents as a series of plots/results—climatology, uncertainty, efficiency, and loads—without a clear discussion connecting the physics behind these results. For example, while the results show efficiency differences linked to BLH, there is limited discussion on the *physical mechanisms* driving these results. The paper would be significantly strengthened by moving beyond a presentation of plots to a cohesive storyline that explicitly links the atmospheric physics (stratification, capping inversion effects) to the observed engineering outcomes (wake recovery, loads).

2. The authors use BLH as a proxy for atmospheric mixing, yet the CFD simulations for efficiency are run under "conventionally neutral conditions" with varying vertical caps. A low BLH in the Baltic Sea (Site C) is typically driven by stable stratification. Stable boundary layers suppress wake recovery much more than neutral layers with a geometric lid. By simulating these conditions as neutral, the study likely underestimates wake losses and overestimates efficiency for the low-BLH cases.

3. It is unclear if the CFD simulations used for efficiency calculations incorporate the large-scale 2D turbulence structures discussed in the second half of the paper. If they do not, there is a disconnection between the two main analyses. The authors should discuss how the efficiency results might change if the wake meandering effects were included in the efficiency loop.

4. It is unclear why 3DTKE was chosen for the WRF simulations (also given that the introduction only talks about the MYNN scheme instead). The authors should briefly justify the choice of 3DTKE for this specific study and perhaps comment on how it handles the capping inversion compared to MYNN.

5. The authors assess WRF BLH uncertainty using Lidar data at Site B (a coastal site) and apply this uncertainty to Sites A and C (offshore/marine sites). The assumption that a coastal site represents a "worst-case" or valid proxy for offshore sites like Dogger Bank (Site A) is strong and potentially flawed. Offshore marine boundary layers are driven by different fluxes than coastal transition zones. If no other data exists (what's the spatial extent of the WRF simulations?), the authors must acknowledge this as a major limitation rather than treating it as a standard conservative assumption.

Specific Comments

- The introduction lacks a critical discussion of previous studies on the topic. It lists research questions but does not sufficiently contextualize the current work against existing literature on BLH-wind farm interactions.
- Figure 5:
  - How was the specific period shown in Figure 5 selected? Was it chosen for its particularly good agreement between lidar and WRF?
  - Is the time axis in UTC?
  - Please define "HWS".
- Section 3.4: When comparing WRF and Lidar BLH, how do the range gate resolution of the lidar and the vertical resolution of the WRF model impact the observed bias? Could the error be an artifact of resolution differences?
- Section 3.5: The load analysis regarding wake meandering appears to be analyzed at a single mean wind speed. Given the paper's emphasis on climatology in the first half, limiting the load analysis to a single wind speed feels restrictive. The authors should expand the discussion on how sensitive these load increases are to wind speed variations (e.g., would the meandering effect diminish at rated power?).

Technical Corrections

- Throughout the manuscript, please add a space between the number and its unit (e.g., "150 m" not "150m") and between unit components (e.g., "m s" not "ms").
- Use exponential notation for units consistently, in accordance with WES standards.
- Tables 2–5: Remove the dot in the unit.
- Table 6: The number of decimals provided is excessive. Suggest reducing them to make comparisons easier.
- Table 7: Please add units for "Average of Bias" and "RMSE".
- Figures 11, 12, 14: Please add where appropriate.
- L. 74: Remove the parentheses around the reference.
- Please add DOIs to all references where appropriate.
- The reference list contains at least one "Discussion" paper that has been published. Please update these to the final published versions.