Overall Assessment

Validating wind farm parameterizations with offshore SCADA data by Sengers et al. demonstrates the value of the recently developed Fitch-pAIM parameterization through a multi-stage validation. The authors first compare the energy yield of Fitch-pAIM against six existing wind farm parameterizations using SCADA observations from six offshore wind farms in a well-studied German Bight region. Based on this initial evaluation, the standard Fitch and EWP schemes are retained for more detailed comparison with Fitch-pAIM at a finer grid cell resolution. Results ultimately suggest Fitch-pAIM's improved turbine power curve representations translate into more accurate energy-yield predictions.

The manuscript is well written and addresses a novel topic of clear interest to the offshore-wind meteorological modeling community. I believe the manuscript would be further strengthened by addressing the following comments prior to publication.

Methodology

1. Defining uncertainty

The manuscript frequently refers to "uncertainty," and only some of these use cases are accompanied by formal uncertainty estimates. I would encourage the authors to distinguish more carefully between “uncertainty”, “variability”, and “sensitivity” and more clearly distinguish what quantity is being evaluated in each case.

2. Atmospheric stability vs. boundary-layer structure

Section 3.4.3 uses model PBLH as "an alternative to any stability metric" to evaluate model performance across different boundary-layer regimes. Can the authors clarify how the binning thresholds were determined and to what degree the modeled NWF PBLH is representative of the observed boundary layer structure? Because boundary-layer depth and atmospheric stability are related but distinct properties, it would also be helpful to discuss this distinction explicitly or, if feasible, corroborate the analysis using an independent stability diagnostic. Such clarification would strengthen the manuscript's conclusion that the parameterizations perform robustly across a range of atmospheric conditions.

3. Generalizability

Throughout the Discussion, the authors identify several factors (e.g., the number and distribution of vertical levels, tuning parameters within the schemes, site characteristics, etc.) that are framed as unlikely to alter the principal conclusions while simultaneously serving as promising avenues for future work. It would be helpful to clarify which aspects of the results are believed to be fundamentally insensitive to these factors, and which may change quantitatively or qualitatively as they are explored in future studies.

Organizational suggestions

1. **Methods versus Results.** Several methodological descriptions are embedded within the Results section. While this organization can sometimes improve readability, expanding the Methods section to introduce the evaluation metrics and their scientific motivation may provide a clearer roadmap for readers here.
2. **Model characteristics versus model performance.** Separating the description of how the parameterizations differ from the subsequent evaluation of their performance may improve the flow.

Editorial Comments

1. L14: "WPFs"
2. L36: Consider "financial" rather than "economical."
3. Figures 1, 8, and 10 could benefit from larger legends, labels, and axis text. The EWP color is also difficult to distinguish in several plots.
4. L60: Perhaps “incompatible” is more appropriate here?
5. L115–129: Consider summarizing this information in a table or incorporating into Figure 1.
6. A summary table of the evaluated model characteristics and performance metrics may help orient readers.
7. L195: "wind speed deficit"
8. L276: Should Section 3.4.3 also be referenced?
9. L472: Consider "generated" rather than "produced."
10. L472: "coarser."
11. L479: “This section”
12. L467: “, as shown in Chang et al. (2026).”
13. L 479: “Its”