
Editor's and reviewers' comments appear in italics; **our responses appear in boldface blue text**.

Reviewer 2:

The study evaluates two planetary boundary layer (PBL) schemes in WRF in the North Sea. Both are compatible with the Fitch wind farm parameterization. For validation, high-resolution measurements of a research aircraft and the FINO1 mast measurements are used. The study finds that the optimal PBL scheme varies with the location, quantity of interest, and error metric.

In contrast to what the title suggests, not only the PBL scheme is varied, but also the TKE advection within WRF is activated and deactivated, and the sensitivity of the results to the TKE factor within the WFP by Fitch et al. is investigated. Due to the many different aspects being studied, as well as the two parameters, different error matrices, and two locations, the results section is rather clouded, and it is difficult to extract the main message. On the other hand, it shows how complex the situation is and that a simple answer to the question of the "optimal" scheme is impossible to give. I think this in itself is an important conclusion and therefore I recommend publishing this article, after some revisions.

We thank the reviewer for their thoughtful assessment and appreciation of the complexity of an "optimal assessment." We believe that the above appropriately characterizes the scope of our analysis.

Major comments

(1) The authors put a lot of effort into investigating this single case study of (mainly) 2 hours (although some FINO1 analysis stretches over the period of 1 day). This imminently raises the question of how transferable those results are to other meteorological conditions (and sites). While the conclusions state that "Subsequent investigations could explore other case studies to provide perspective into the generalizability of the results across other sites." the author also critically highlights from previous literature that "Conclusions drawn from these validation studies may also be influenced by site-specific or meteorological conditions." Thus, the discussion on the generalizability should be extended to highlight why it is worth spending so much effort on just 2 hours.

We thank the reviewer for their thoughtful comment. Reviewer 1 similarly addressed the duration of the analysis. The 2 hours of aircraft data aligned with the 12 hours of FINO1 analysis were strategically chosen to accomplish several goals, including:

- **Contextualizing our results with the broader literature for this case study (i.e. consistency with Ali et al. (2023)**
- **Providing some level of indication as to the diurnal consistency of our results. For example, the optimal PBL scheme in the FINO1 region is consistent throughout the time period of the study**
- **Restricting our analysis to high-quality, in situ observations without introducing additional uncertainties associated with methods like statistical downscaling**
- **Restricting our analysis to stably-stratified conditions, which contribute the strongest and longest wakes.**

Thus, by restricting our analysis to conditions that are well-suited for wake analysis, with in situ observations that reduce uncertainty, and also also allow us to contextualize our results with the broader literature for this case study, we believe our analysis reframes the apparent lack of available long term in situ offshore wind farm observations into a well-suited case study analysis for documenting physical differences between models.

We also now include a similar justification in our discussion section in (updated) lines 531-536:

"The stable stratification present in this case study also improves the utility of the results of this PBL comparison. By restricting this analysis to time periods considered in previous analyses for this case study, not only are the results contextualized within the broader literature, but the conditions that contribute the strongest and longest wakes are also highlighted. Thus, while other analyses of this region may approach the lack of available in situ observations by introducing statistical downscaling methods to explore scientific questions around diurnal, seasonal, and climatic trends (Fischereit et al., 2022b), this analysis instead addresses scientific questions that are best-suited with in situ observations alone."

(2) The study mainly compares the two PBL schemes, the 3DPBL and the MYNN scheme, in a wind farm context. Thus, along with a short introduction to the WFP by Fitch et al., it should also provide some introduction to the two schemes and the difference between those to extend the too brief description that is given in the introduction (line 72-81). This will also aid the interpretation of the results.

We thank the reviewer for this thoughtful comment. We have added a full subsection to our methods section (the new Sect. 2.1) that includes the governing equations for both models and discusses the key differences between the PBL schemes. We then reference this section throughout the manuscript to aid interpreting the physical differences between the schemes. As some examples:

• L. 351: “MYNN simulations … the surface.”

In this section, now starting on line 351, we argue that the slight differences in stability are a consequence of the increased TKE from the 3DPBL scheme. This increased TKE with the 3DPBL scheme encourages slightly more mixing:

“Because the modeled air temperatures are almost identical between models (Fig. 4a,f), these slight differences in the surface stability could be a consequence of the greater TKE with the 3DPBL scheme that encourages slightly more mixing”

• L. 390: “Both the … simulation (Fig. 6a,c).”

In this section, now starting on line 390, we point out that the larger TKE with the 3DPBL scheme in stable conditions, which is consistent with the trend identified under the idealized, stable conditions simulated in Rybchuk et al. (2022), reflects the fundamental differences between the 3DPBL and MYNN models:

“These differences in TKE between the PBL schemes, which are consistent with those identified under the idealized, stable conditions simulated in Rybchuk et al. (2022), reflect the fundamental differences between the models. Notably, while the MYNN scheme uses Smagorinsky mixing to characterize horizontal turbulent mixing, the 3DPBL scheme instead calculates the horizontal turbulent flux divergences explicitly. The two models also rely on different length scales and empirical constants”

• L. 417 “MYNN average … wind speeds.”

In this section, now starting on line 417, we maintain our interpretation that larger TKE with the 3DPBL scheme implies greater momentum extraction that results in reduced wind speeds.

“This larger TKE with the 3DPBL scheme extracts more momentum from the mean wind, resulting in a greater reduction in wind speed. This finding that MYNN wind speeds are faster than 3DPBL wind speeds is consistent with other comparisons of these two PBL schemes, completed in both real and idealized conditions (Juliano et al., 2022; Rybchuk et al., 2022; Arthur et al., 2022; Peña et al., 2023; Arthur et al., 2024).”

We also introduce a new Fig. 7 to document the differences in vertical structure of wind speeds between the two PBL schemes to support our claim.

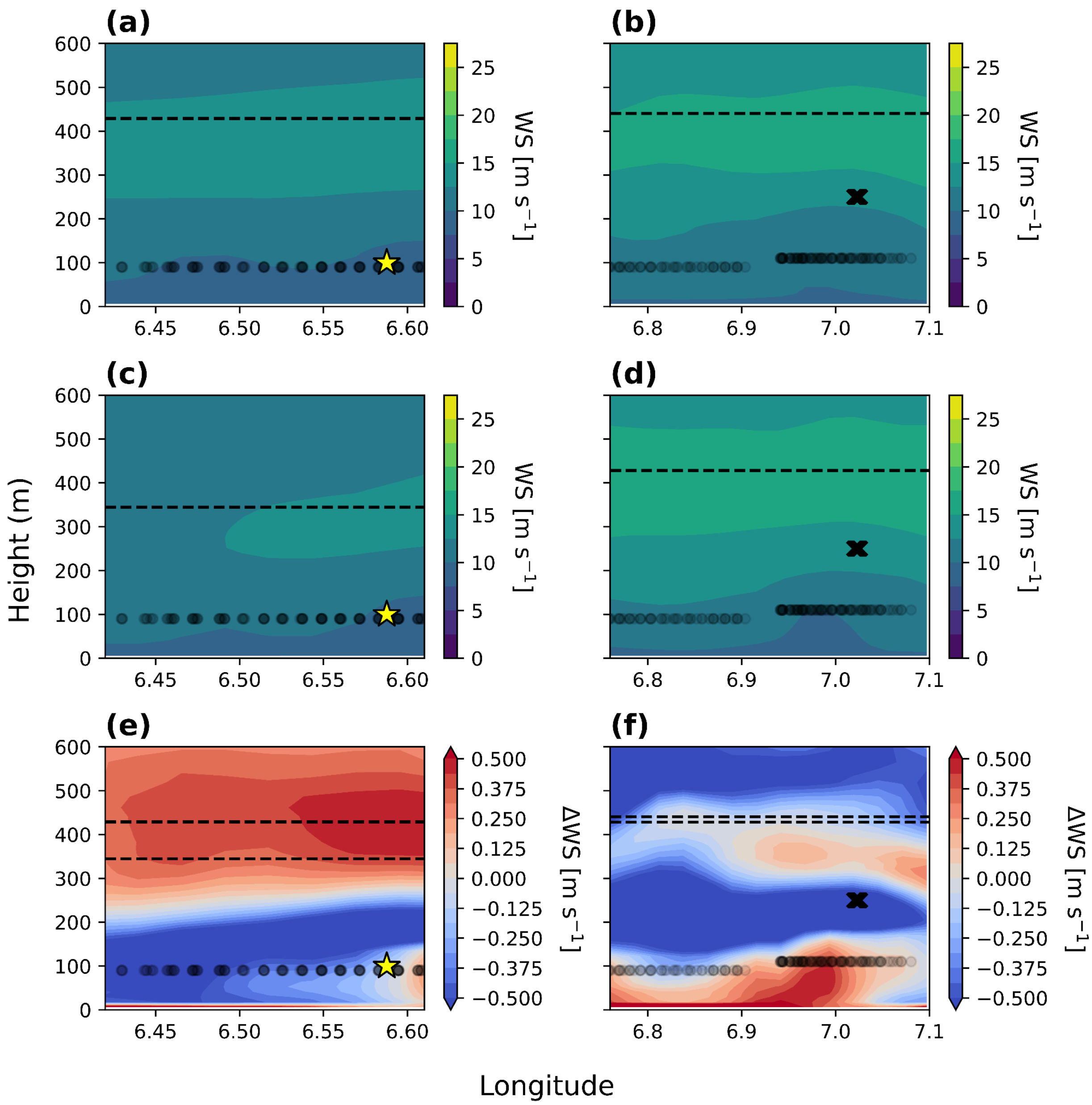

Figure 7. Modeled wind speed cross-section at a constant latitude of 54.03. (a) FINO1 3fa_025; FINO1 mfa_025; (e) FINO1 3fa_025- mfa_025; (b) aircraft 3fa_025; (d) aircraft mfa_025; (f) aircraft 3fa_025 - mfa_025. The horizontal dashed black line denotes the average modeled PBL height, the star indicates the FINO1 tower location, the "X" marks the first transect path, and the black circles indicate the turbine hub height.

Finally, we corroborate this interpretation in Sect. A2 of the Appendix. Notably, Fig. A3a of Sect. A2 demonstrates how increasing the amount of TKE (in this case, via the wind farm TKE factor) reduces the wind speeds.

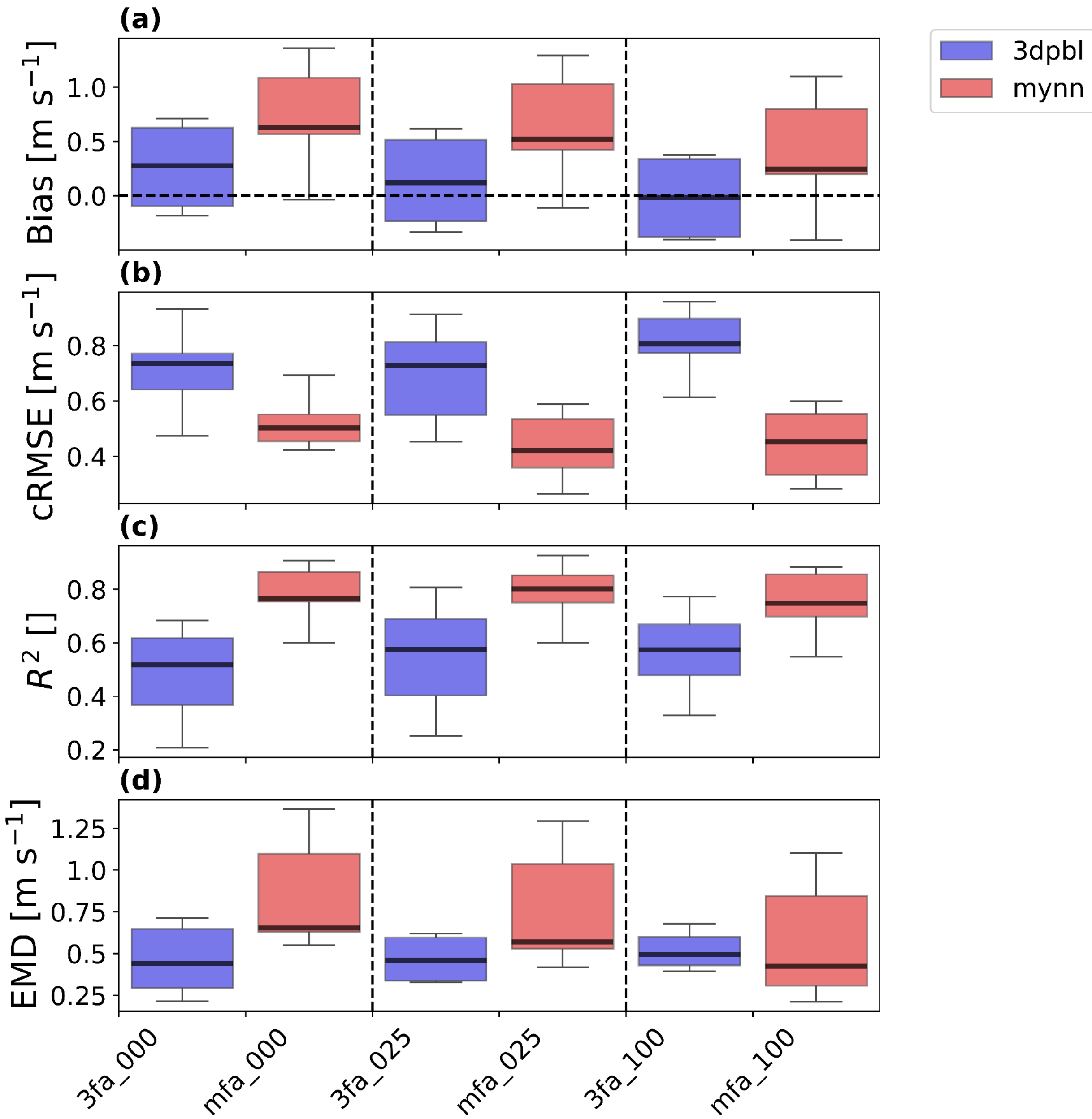

Figure A3. Error metric box plot for aircraft observations collected at 250 m for wind speed. The box and whiskers describe aircraft transect variability and are based on Q1 (25th percentile), Q3 (75th percentile), and the interquartile range (IQR) (Q3–Q1). The box encloses the IQR, and the whiskers extend to Q1-1.5*IQR and Q3+1.5*IQR. The simulation names are mapped according to the short names provided in Table 6, and the vertical dotted lines visually separate simulations by wind farm TKE factor. (a) Wind speed bias; (b) wind speed cRMSE; (c) wind speed R2; (d) wind speed EMD.

- **L. 449 "TKE in … MYNN simulations."**

In this section, now starting on line 449, we reference the figure from the discussion in the previous bullet point and maintain the same interpretation.

“As noted earlier, these differences in TKE between the PBL schemes reflect the fundamental differences between the models in length scales, empirical constants, and horizontal mixing approaches.”

(3) Validation of atmospheric stability: Stability is an important parameter for analyzing the results. However, while it is discussed thoroughly from the simulations, the profiles are not validated with the measurements in the lower part of the atmosphere. Both the FINO measurements and the profiles flown by the aircraft could be used to evaluate the simulations in the lower part of the PBL.

We thank the reviewer for their thoughtful guidance. We now include vertical profiles both at the FINO1 tower and from the aircraft profile flights to demonstrate the stable stratification in the new Fig. 4 and Fig. 5. The stable profiles that we now provide in this manuscript are also corroborated by those in Larsén & Fischereit(2021) and Ali et al. (2023).

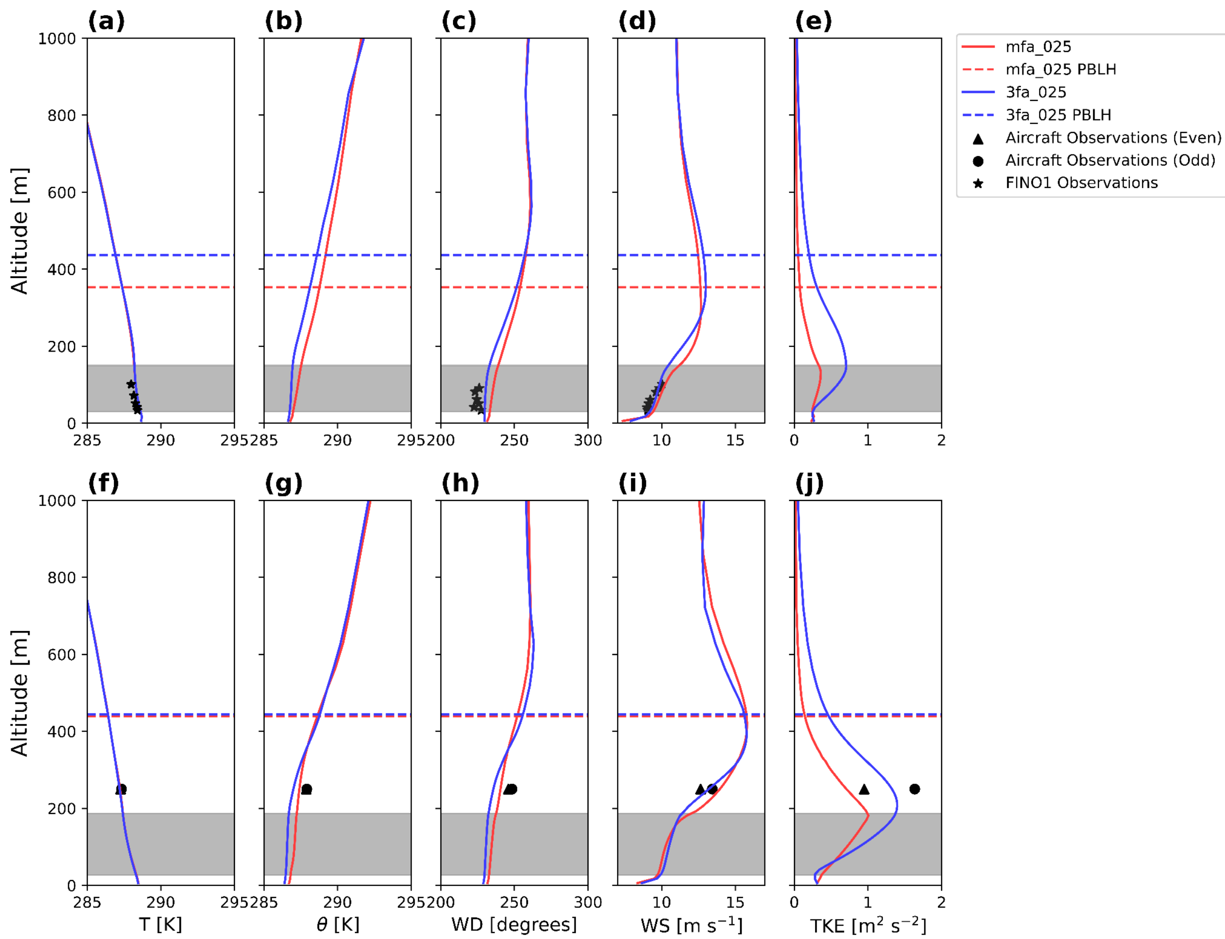

Figure 4. Temperature, potential temperature, wind direction, wind speed, and TKE vertical profiles from observations and WRF simulations for both sites. In all cases, the dashed lines indicate the modeled PBL height and the grey region indicates the turbine rotor region. Observations in the aircraft region are separated between and even and odd transects. (a) FINO1 temperature; (b) FINO1 potential temperature; (c) FINO1 wind direction; (d) FINO1 horizontal wind speed; (e) FINO1 TKE; (f) aircraft temperature; (g) aircraft potential temperature; (h) aircraft wind direction; (i) aircraft horizontal wind speed; (j) aircraft TKE. FINO1 cases are averaged over hours 12:00:00–00:00:00 and the aircraft region cases are averaged over 14:10:00–16:10:00. FINO1 TKE calculations based on observations were not available due to the coarse temporal resolution of the wind speeds. FINO1 potential temperature calculations based on observations were not available due to a lack of pressure observations.

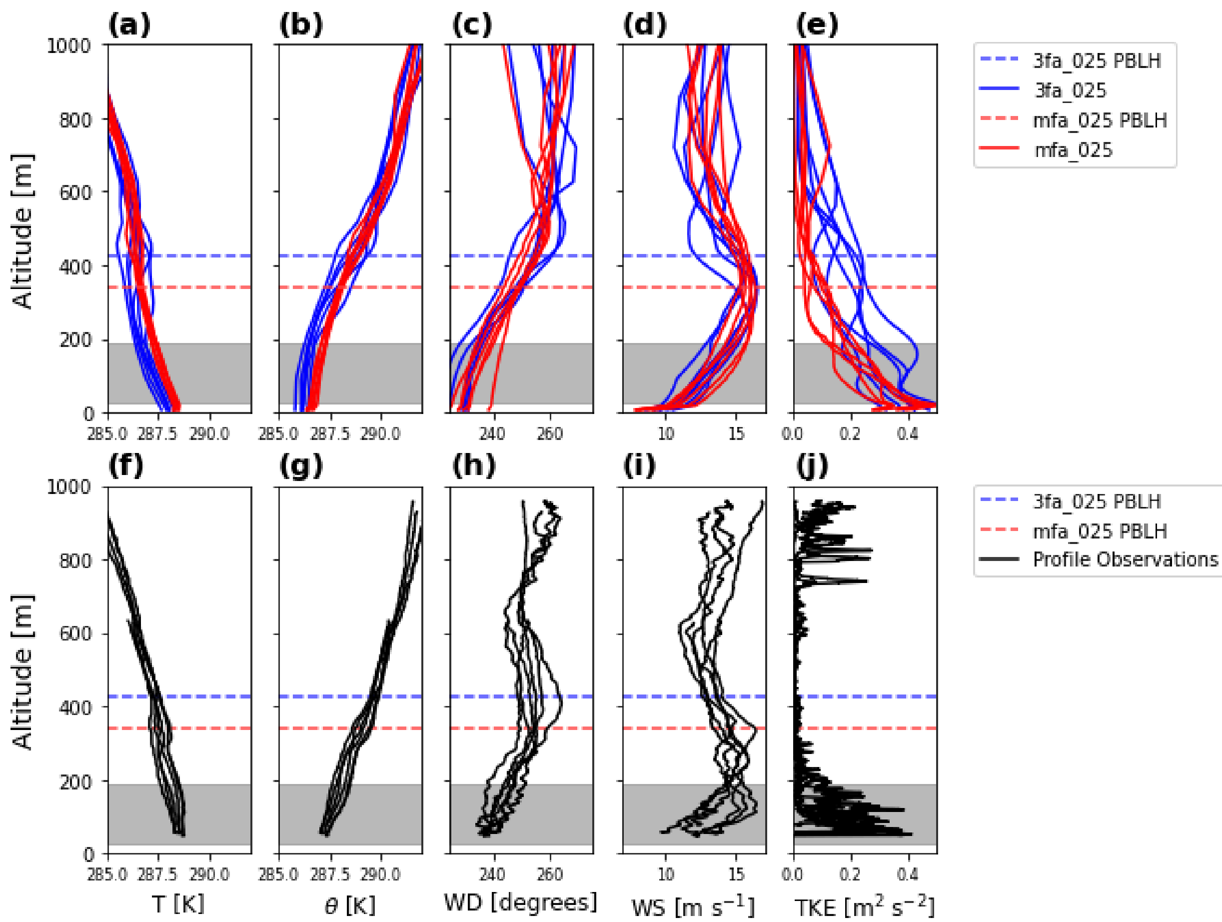

Figure 5. Observed and modeled vertical profiles for the aircraft vertical profile flights (Table 3, Fig. 1b). In all cases, the horizontal line indicates the modeled PBL height, and the color differentiates the PBL scheme. The top row of panels corresponds to modeled output and the bottom row of panels corresponds to the aircraft profile observations. Modeled output are determined to be a given middle cell for each profile as in Larsén and Fischereit (2021) based on the timestep indicated in Table 3. (a) modeled temperature; (b) modeled potential temperature; (c)
modeled wind direction; (d) modeled horizontal wind speed; (e) modeled TKE; (f) observed temperature; (g) observed potential temperature; (h) observed wind direction; (i) observed horizontal wind speed; (j) observed TKE.

(4) The appendix is too long and not only adds supporting information but presents new findings, e.g., regarding the effect of TKE advection. I suggest moving the section on the TKE advection to the main text, while some plots can probably be kept in the appendix (or removed completely) to limit the size of the main text. Furthermore, some additional figures can be removed if they do not add any new information. This would make the Appendix much more accessible, as now it is flooded with figures.

We thank the reviewer for this thoughtful guidance, which is also echoed by Reviewer 1. Based on this joint guidance, we have revised the Appendix in the following method to make it more of a story by doing the following:

- **We have removed the entirety of Appendix B from the text**
- **We have preserved the NWF vs. WF distinction as part of the appendix to better distinguish the differences between PBL schemes in "baseline" conditions (Sect A1)**
- **We have moved the discussion of the effect of the wind farm TKE factor to the Appendix (Sect A2)**
- **We have kept the advection section in the Appendix (now Sect A3) to focus the manuscript. We considered the reviewer's suggestion to move advection into the main text but we feel that would unnecessarily broaden the discussion in the main paper.**

Specific comments

- Line 47: Other (newer) references should be mentioned

We thank the reviewer for this comment. Reviewer 1 similarly acknowledged that a citation from 2014 was not necessarily representative of the most current advancements in wind farm parameterization validations. We have updated this reference to a more recent, highly-cited review of wind farm parameterization validations (Fischereit et al., 2022).

- Line 59: Add citation for EWP

Thank you, we have added the citation for EWP on (updated) line 62.

- Table 1: "Select" -> "Selected"

Thank you, we have adjusted the Table 1 caption to show "selected" instead of "Select".

- Section 2.2 Provide a tabular overview of the WRF settings to provide a better overview for the readers. Also, consider publishing the namelist settings on Zenodo for reproducibility.

Thank you, we have added a table of select WRF settings (Table 5) and also intend to publish our namelist settings on Zenodo upon acceptance of the paper as noted in the Data Availability section.

- Line 131: Which WRF version is used precisely?

We have added a reference to WRFV4.4.2 on line 206.

- Line 157: Consider repeating, "The results from the other runs are analyzed in the Appendix."

Thank you, we repeatedly emphasize that the results of the other runs are analyzed in the Appendix on (updated) lines 224, 240, and 254.

- Line 205: Consider removing "also"

Thank you, we have removed “also” on updated line 319.

- Figure 4: Why are not all curves labeled? This might be useful for other researchers.

Thank you, we have added labels and unique colors to each of the drag proxy curves for what is now Fig. 3.

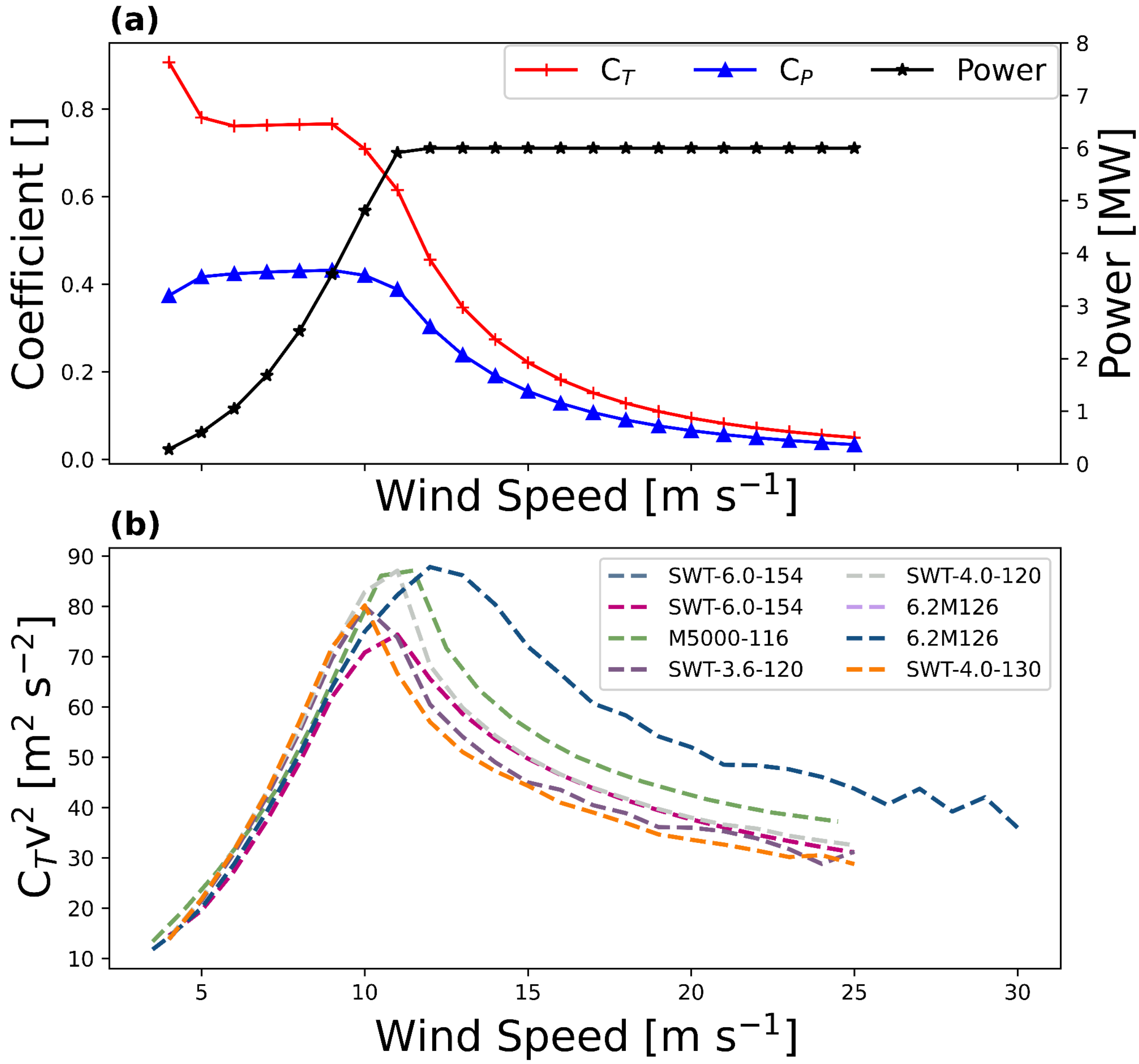

Figure 3. (a) Curve illustrating turbine C_T, C_P, and power specifications for the turbine model in the Gode wind farm. (b) Drag proxy for each of the eight turbine models present in this case study.

- Figure 8: Why is the Mean (solid line) sometimes above the single realizations (as shown by the symbols)? Or does "Mean" mean something else here?

Thank you, we acknowledge the lack of clarity in this figure (now Fig. 9). Reviewer 1 has similarly expressed confusion about this plot. By focusing on only one wind farm TKE factor, we hope that this figure is now easier to interpret.

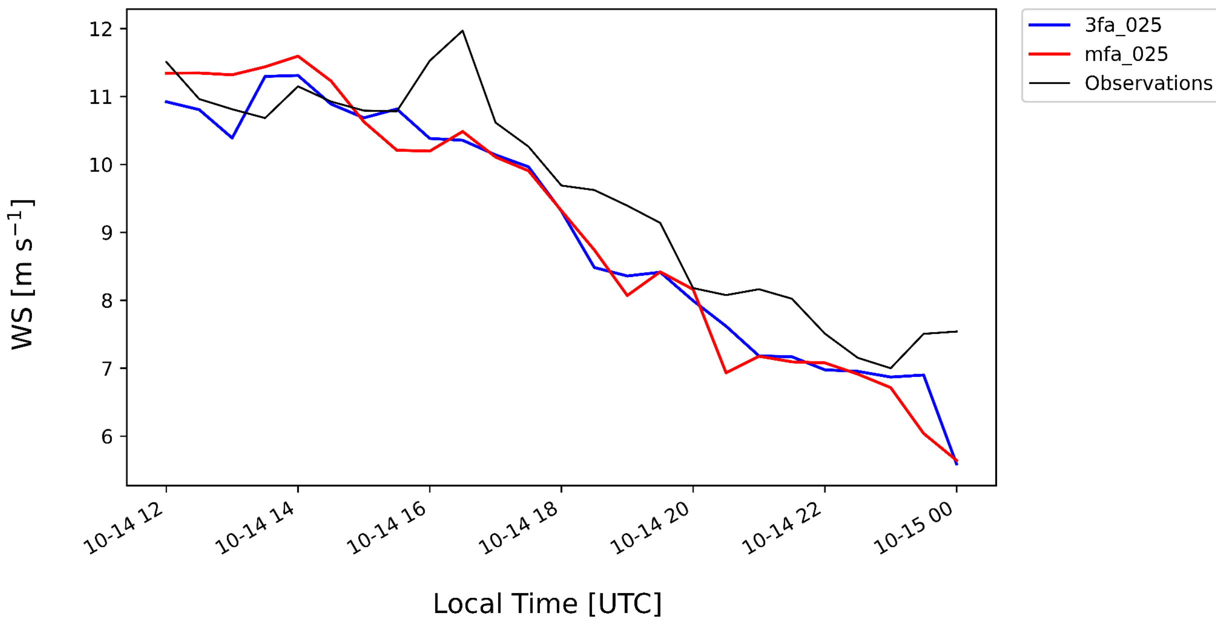

Figure 9. Time series of 76 m modeled horizontal wind speeds (WS) compared to 81 m FINO1 observations for the hours of 12:00:00–00:00:00. Both the modeled wind speeds and observed wind speeds are resampled to 30 minutes.

- Figure 14 / 15 captions: Put the analyzed quantity at the beginning of the caption to easier see the difference between those two figures: Error metric box plot for aircraft observations collected at 250 m for wind speed / TKE.

Thank you, we have adjusted our captions to read as "Error metric box plot for aircraft observations collected at 250 m for wind speed / TKE." These figures are now listed as Fig. A3 and Fig. A4.

- Line 408: add space between "(2023)," and "Optis"

Thank you, we agree that there should have been a space. However, this sentence is now removed as part of the removal of the discussion of the optimal wind farm TKE factor.

- Figure A1 caption: Use also abbreviations as documented in table 4 in the caption

Thank you, we have introduced the Table 4 abbreviations in the caption for Fig. A1.

- Figure B10-B13: See comment regarding figure 14/15

Thank you, this figure has been removed entirely, consistent with the distilling of the Appendix.

- Line 632: "..wind speeds - to - outperform ... " add "to" here?

Thank you, we agree. This sentence has been removed entirely, consistent with the distilling of the Appendix.

References

Ali, K., Schultz, D. M., Revell, A., Stallard, T., & Ouro, P. (2023). Assessment of five wind-farm parameterizations in the Weather Research and Forecasting model: A case study of wind farms in the North Sea. *Monthly Weather Review*, *1*(aop). https://doi.org/10.1175/MWR-D-23-0006.1

Bärfuss, K. B., Schmithüsen, H., & Lampert, A. (2023). Drone-based meteorological observations up to the tropopause – a concept study. *Atmospheric Measurement Techniques*, *16*(15), 3739–3765. https://doi.org/10.5194/amt-16-3739-2023

Fischereit, J., Brown, R., Larsén, X. G., Badger, J., & Hawkes, G. (2022). Review of Mesoscale Wind-Farm Parametrizations and Their Applications. *Boundary-Layer Meteorology*, *182*(2), 175–224. https://doi.org/10.1007/s10546-021-00652-y

Larsén, X. G., & Fischereit, J. (2021). A case study of wind farm effects using two wake parameterizations in the Weather Research and Forecasting (WRF) model (V3.7.1) in the presence of low-level jets. *Geoscientific Model Development*, *14*(6), 3141–3158. https://doi.org/10.5194/gmd-14-3141-2021

Rybchuk, A., Juliano, T. W., Lundquist, J. K., Rosencrans, D., Bodini, N., & Optis, M. (2022). The sensitivity of the Fitch wind farm parameterization to a three-dimensional planetary boundary layer scheme. *Wind Energy Science*, *7*(5), 2085–2098. https://doi.org/10.5194/wes-7-2085-2022