We thank both the editor and reviewer for considering this manuscript. Below, reviewer comments are in black and **our responses are in blue bold.**

Third round review of "A North Sea in situ evaluation of the Fitch Wind Farm Parametrization within the Mellor–Yamada–Nakanishi–Niino and 3D Planetary Boundary Layer schemes" by Agarwal et al.

The manuscript has undergone significant change during all revisions. After revising some minor comments listed below, the manuscript will be ready for publication.

Specific comments
L 374-376. Here it is stated that since 3DPBL has a higher wind speed in the wake than MYNN, it also has more turbine-induced TKE. However, the turbine-induced TKE is only related to the TKE source from Fitch, which actually should be lower in 3DPBL than in MYNN because of the lower wind speed. Rather, 3DPBL generates more mechanical (shear-driven) turbulence. This should be noted and the physical mechanism (ideally refer to 3DPBL equations) should be explained.

We thank the reviewer for the close concern for physical mechanisms. We also wish to acknowledge that the above does not accurately reflect the PBL-based differences in wake behavior. While the reviewer accurately corroborates our physics-based argument for the underlying meteorology, an additional mechanism is also at play in the wakes. Starting at the beginning of the noted paragraph on line 367, we explain:

"Once these faster winds, in the MYNN scheme, enter the region of a wind farm, the faster winds lead to a larger drag force exerted by the wind turbines (Fig. 3b) and therefore a larger wake effect, resulting in slower winds in the MYNN wakes (Fig. 8a). This distinct behavior in the wakes arises from differences in the drag forces for each PBL scheme. The drag forces are very sensitive to wind speed (Fig. 3b). Because the MYNN wind speeds are slightly faster when entering the wind farms, the resulting MYNN drag force (Eq. 11, Fig. 3b) is generally larger than the 3DPBL drag force. As a consequence, the MYNN scheme shows stronger and longer wakes than the 3DPBL scheme, on average (Fig. 8c). The MYNN average wind speed reduction is sufficiently strong such that 3DPBL average wind speeds exceed MYNN average wind speeds within the turbine wake (Fig. 8a). Further, because 3DPBL average wind speeds exceed MYNN average wind speeds in this region, the 3DPBL scheme also has more turbine-induced TKE than the MYNN scheme (Fig. 8b). This turbine-induced TKE can help erode the wake in the 3DPBL simulations."

We also corroborate this wake mechanism discussion in a different figure and accompanying discussion. Starting on line 534 of Appendix A1:

"…Fitch and NWF wind speeds differ near the turbines. Notably, 3DPBL Fitch average wind speeds exceed MYNN Fitch average wind speeds in the turbine wakes (Fig. A1f). This reversal of which PBL scheme shows the faster average wind speed can be explained by differences in the turbine drag force between the two PBL schemes.

The MYNN scheme has a stronger turbine drag force (Eq. 11, Fig. 3b) because of its faster initial wind speeds, which also implies that the MYNN scheme has stronger and deeper wakes (Fig. 8c). This reversal (of which PBL scheme shows the faster average wind speed) occurs only in the

monotonically increasing region of the drag (proxy) curve (Fig. 3b). If the wind speeds were instead within the monotonically decreasing region of the drag (proxy) curve (Fig. 3b), MYNN wind speeds would likely exceed 3DPBL wind speeds even in the wakes because the faster winds would result in a smaller drag force (Fig. 3b). Given that NWF wind speeds mirror Fitch wind speeds outside of the turbine wakes and NWF wind speeds differ from Fitch wind speeds within the wakes, the dominant mechanism for these differences is more likely related to the turbines and not to the underlying meteorology."

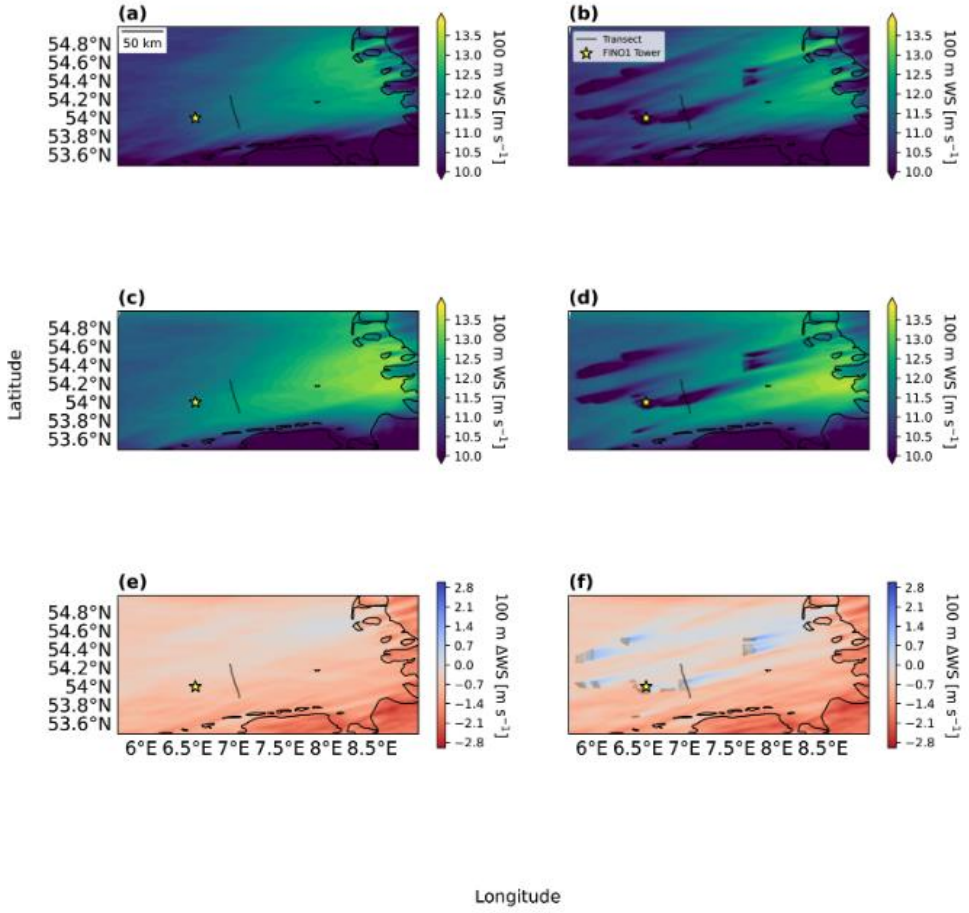

Figure A1. Horizontal wind speeds for the inner 1.67 km domain. (a) 3na_NA. (b) 3fa_025. (c) mna_NA. (d) mfa_025. (e) 3na_NA - mna_NA. (f) 3fa_025 - mfa_025. Turbines are marked with black circles, the FINO1 tower is marked with a yellow star, and the first transect path is marked with a solid line.

Appendix A4: Thanks for adding the 28 day analysis. Your current final sentence "Increasing the length of the evaluation window helps reveal systematic differences between the model" is a rather weak and informative statement. Please add more details on what we can learn from this analysis. With the mention of 3 statistically significant results, you should add that bias is significantly lower for MYNN, while cRMSE and R2 are significantly better for 3DPBL. Most importantly, explicitly mention whether extended analysis confirms the takeaways from the 12h analysis or not. Lastly, please use the same labels on the x-axis as in Fig 14 and in the main text directly refer to this appendix to make the reader aware.

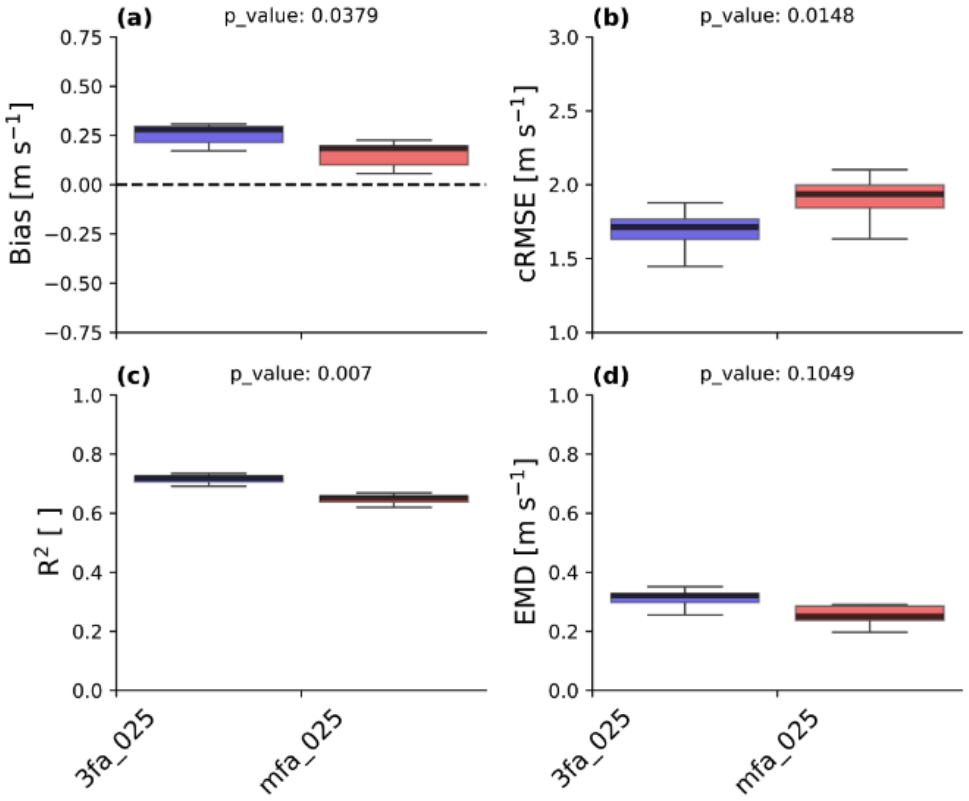

Figure A7. Error metrics across all Fitch simulations with the advection option on and with a wind farm TKE factor of 0.25 at the FINO1 site for the period of Oct 1 - Oct 28 for wind speed. The bar shows the median. The box encloses the interquartile range (IQR), and the whiskers extend to Q1-1.5*IQR and Q3+1.5*IQR. (a) wind speed bias; (b) wind speed $cRMSE$; (c) wind speed R^2; (d) wind speed EMD.

Thank you. We offer the following updated analysis of the label-refined figure starting on line 625:

"The PBL schemes show mixed performance in the extended FINO1 evaluation. 3DPBL produces lower median cRMSE (Fig. A7b) and higher R2 (Fig. A7c) than MYNN, whereas MYNN produces lower bias (Fig. A7a) and lower EM D (Fig. A7d) than 3DPBL. However, only the differences in cRMSE (Fig. A7b), bias (Fig. A7a), and R2 (Fig. A7c) are statistically significant; the EMD (Fig. A7d) difference is not. Considering only statistically significant error metrics, 3DPBL performs better for two of the three metrics.

The longer time period also clarifies model differences. In both cases, 3DPBL shows the optimal cRMSE. However, in the twelve hour analysis, statistically significant differences only emerged for cRMSE (Fig. 14d), whereas the 28 day analysis instead suggests statistically significant differences

with respect to three error metrics (Fig. A7). Ultimately, the 28 day analysis corroborates the twelve hour analysis by more fully confirming 3DPBL as the optimal scheme for this location with respect to more error metrics."

Technical corrections
L8 "3DPBL TKE bias underperforms MYNN TKE bias": not very clear what underperforming here means. Please rephrase
This line now reads "3DPBL TKE bias is larger than MYNN TKE bias…"

Section 2.1.3: for me not necessary, but ok to keep it in.
We preserve this section out of consideration for other readers who have expressed interest in this information.

Section 3.1.1 header: not sure whether "atmospheric stability" fully captures the content of this section. Proposal: "Profiles of atmospheric variables" or similar
We offer "Profiles of atmospheric variables" as a title.

L360: 3PBL ->3DPBL
The typo has been addressed.

L365: In this sentence you talk about both TKE (Fig b) and WS (Fig a), yet only refer to Fig b. Either add reference to Fig a or remove this reference to Fig b.
We add the reference a) to the figure.

Fig 11: quality of figures needs to be improved
The quality of Fig 11 has been updated to 1200 DPI.

Conclusions: rename to 'Conclusions and future work'. I'd make this explicit as this comprises a significant part of this section. Please also add that power generation estimates need to be evaluated in future work.
We alter the section header to "Conclusions and future work". We also have added an additional final paragraph to this section:

"Power generation analyses for this case study are another future work opportunity. Analyses like Mittelmeier et al. (2017) and Sanchez-Gomez et al. (2023) have demonstrated the utility of the available North Sea SCADA data in informing cluster wake research. As such, an accompanying analysis that validates the power estimates derived with each PBL scheme for this region may offer additional insight into model differences and draw interest from a broader set of stakeholders."

L495 "The optimal PBL scheme depends on the site": add that it also depends on the variable and error metric
This line now reads "the optimal PBL scheme depends on the site, variable, and error metric."