We thank the referees for their careful reading of the revised manuscript and for their recommendations. Please find our responses below in blue. Note that line numbers refer to the main pdf, not the version with changes tracked.

Response to Referee 1

Introduction and recommendation

This manuscript compares simulations of atmospheric flow at a wind power production site in central California, performed with the WRF model at 1-km horizontal grid spacing and two alternative parameterizations for planetary boundary layer (PBL) turbulence. My first review of the manuscript highlighted some deficiencies in the interpretation of forecast verification statistics (systematic vs. random errors). I also recommended de-emphasizing the discussion of the differences between the two PBL schemes, which appeared to be rather marginal in practice. I am generally happy with the revisions, with the single exception of the answer to my minor comment 10, which I do not find entirely satisfactory. I appreciated the addition of Appendix A and Figures A1 and A2, which show the impact of the PBL schemes more convincingly.
I recommend acceptance, provided that lines 260-265 and 290-293 of the revised manuscript are edited as described below.

Thank you, we appreciate your suggestions on both the original manuscript and the revision.

Specific (minor) comments

1. One of my comments on the first version of the manuscript (number 10) was about the negative wind speed bias at z < 30 m AGL, visible in Figure 3b. At lines 260-265 of the revised manuscript, the authors interpret the negative bias during speedup events (18-21 LST) using mass continuity arguments:

"Conversely, wind speeds are underestimated near the surface, indicating that the model fails to capture near-surface accelerations. ... While the model captures some negative vertical velocities at the study site during the speedup events (see contours in Figure 3f), its vertical velocities are too weak and thus do not translate to near-surface accelerations of the magnitude seen in the observations."

This argument might explain why the negative bias becomes slightly larger during the evening speedup events, but not why it persists throughout the whole diurnal cycle. A remark was added in response to my comment, at lines 290-293 of the revised manuscript:

"This may be due to slightly improved predictions of vertical mixing of higher momentum from aloft; during this time, prior to jet development, the winds follow a standard quasi-logarithmic profile. The 3D PBL scheme has been shown previously, in idealized tests, to improve model performance during daytime convective conditions (Juliano et al., 2022)".

This explanation misses the point, because it refers to the small difference in bias between the two simulations with different PBL schemes; not to the bias itself, which remains quite large in both cases. A better explanation of the persistent negative wind speed bias might be that "predictions of vertical mixing of higher momentum from aloft" are quite bad even with the 3D-PBL scheme (Figure 3b), which however does a marginally better job than the competitor (Figure 4).

Thank you for clarifying this point. We agree that inadequate vertical mixing is an important component of the model performance in both PBL configurations. To address your comment, we have decided to keep the above lines as is, but to work the suggested point into the existing higher-level discussion on lines 307-326. In particular, lines 307-313 now read as follows:

"Taken together, wind speed error metrics (Table 2), composite-average profiles (Figures 3 and 4), and results from the sample day (Figure A1) suggest that for both model configurations, the predicted amount of vertical mixing is inadequate to transport higher momentum downward from aloft. This results in a persistent negative wind speed bias below roughly 30 m AGL throughout the day. During speedup events, too much momentum remains within the rotor layer. Although both model configurations produce a pronounced jet below hub-height and reduced wind speeds above (Figure 4, 2100-0600 PST), wind speeds are generally overestimated in the rotor layer and underestimated near the surface."

The issue is not particularly relevant in the wind energy context, because the near-surface layers are essentially irrelevant for wind power harvesting. It might be important in other contexts, though.

We hope that our study is of interest to a general atmospheric modeling audience, so we appreciate this consideration.

Response to Referee 2

General considerations

The authors have adequately addressed the issues raised by the reviewers in the first round. In particular, the error metrics are now correctly defined and the claimed improvement due to the 3D PBL scheme is more carefully (and hence more appropriately) formulated. I was pleased to see the introduction of the case study day (in the appendix) and I think that also the introduction of at least two layers for the determination of error statistics has produced some added value. However, going through the manuscript once more (or for the first time, for the appendix), I detected some issues that still need to be addressed, before the paper can be published. For formal reasons, I have labelled the one concerning the new appendix as 'major' – but as it is straight forward to be addressed, I don't think that it requires an additional review round.

Thank you, we appreciate your suggestions on both the original manuscript and the revision.

Major comment

1) New appendix A: this is in fact a very interesting (and also uncovering) additional part of the manuscript. I think it clearly demonstrates the consequences of heavy averaging (in the vertical),

e.g. leading to the numbers in Tab. 2 – even if it is now distinguished between two 'layers'. In particular, it seems to demonstrate that both parameterizations have too much vertical mixing, producing a deep vertical maximum of some 50 m depths, while the observations show this 'speedup layer' to be about half as deep. While I think that the overall 'conclusion' ('..highlighting potential model improvements…') is rather optimistic, I only comment on the obvious error: in Fig. A1 the lowest observation is at zero m AGL, thus suggesting that at times there is a (mean hourly) wind speed of >6 m/s at the ground. This must be corrected to 10 m (first range gate of the lidar). This also affects Fig. 4 (again, I have to say that I overlooked this in the first review), and possibly Fig. 3a, c.

Thank you for catching this subtle point that was not fully explained in the manuscript. Figures 3, 4, and A1 include observed wind speed and direction data from the lidar's on-board meteorological station at 1 m AGL. These data are shown to provide additional context for the shape of the near-surface jet. However, model errors are not evaluated at this height because extrapolation below the lowest model grid point (i.e., the lowest half level at which velocities are calculated) would be required. We now explain this on lines 240-242 when introducing the error metrics:

"Although several figures herein present observed wind speed and direction from the lidar's on-board meteorological station at 1 m AGL, model errors are not evaluated at this height because extrapolation below the first half level (at roughly 8 m AGL) would be required."

An abbreviated note has also been added to the three figure captions.

Detailed comments

Fig. 3d I overlooked in the first review that wind direction bias (panel d) is given in m/s instead of °.

This has been corrected.

Tab. 2, caption (and l. 283): why is the 30 m level included in both averaging ranges? As it corresponds to the given definition of the 'rotor layer', it should probably be removed from what is called the 'surface layer', which does not correspond to the traditional surface layer – lowest 10% of the BL height – but possibly to the Inner Layer in the linearized theory of Hunt and colleagues (see the summary paper of Belcher and Hunt (1998) for the references).

We have corrected the vertical averaging ranges for Table 2 such that the 30 m level is only included in the rotor layer. As seen in the diffed pdf, this correction results in minor changes to some of the near-surface metrics, which tend to accentuate the differences between the two layers. We have also generalized the wording to prevent confusion with the layers defined by boundary-layer theory, in both the Table 2 caption and in the text on lines 283-286:

"The metrics shown in the table are time averaged over the full study period and vertically averaged over two separate layers: the rotor layer (lidar measurement heights of 30, 38, 50, 60, 70, 80, 90, and 120 m AGL) and a near-surface layer below the rotor layer (lidar measurement heights of 10 and 20 m AGL)."

References

Belcher S.E., Hunt J.C.R. (1998) TURBULENT FLOW OVER HILLS AND WAVES. Annual Review of Fluid Mechanics 30, 507-538, doi: 10.1146/annurev.fluid.30.1.507.