Dear authors,

Thanks a lot for your revised version. I can see you have undertaken an extensive revision of the manuscript and shorten it. I understand the main author is a PhD student and he is obviously dealing with a complex subject. But the manuscript lacks focus and having both idealized and real simulations together does not help. As the other reviewer and I mentioned in the previous review, the real simulations do not help the manuscript, the understanding of the physics/parametrizations, and makes it long. The real simulations need to be erased from the manuscript in order to move further. They do not need to be wasted; the PhD student can use them in another manuscript (if appropriate) or in his thesis dissertation. In the following, you find a list of major and minor/specific comments.

Dear reviewer, thank you for your continued thoughtful feedback on our manuscript. We have updated our manuscript accordingly. We will also note that as the review process has been going on, we have found and fixed the bug regarding strange near-surface TKE values in the idealized 3DPBL simulations. Thankfully, as we anticipated and noted earlier in the review, this bug had negligible effects on wind speed and TKE profiles.

Major comments:

1. Real simulations: There is no need for them. They do not clarify the role of the PBL scheme on the results. They do not clarify the role of the parametrization on the results either. Having them makes the manuscript long, and unfocused.

Following feedback from Reviewer 1 as well as the editor, we have entirely struck the presentation of the real simulations from the manuscript. However, following suggestions from Reviewer 1, we cite the lead author's dissertation and briefly discuss the results of the real simulations in the Conclusions (L360).

2. If the real-simulations are dropped then you can perform the idealized simulations with the full 3DPBL model, instead of using the "PBL-approximation". This will be definitively interesting when comparing the PBL schemes.

While we agree that analysis with the full 3DPBL model would be interesting, that analysis is beyond the scope of this manuscript.

3. Why are the idealized simulations setup with such high domains (20 km)? This is probably the reason why your simulations take so long to spin up. Why not using 2 to 5 km?

In our idealized simulations, we do our best to replicate the original Fitch et al. (2012) WFP study. That study had its model top set at 20 km (page 3021 of the paper), and as such, we use the same height.

4. Figure 3: Why is for the stable simulation TKE not linearly decreasing with height for the 3DPBL compared to MYNN? Why are the rotor height TKE values so different between the two PBL schemes under unstable and stable conditions? Why is the unstable TKE so different and low in MYNN compared to 3DPBL (it is even half the value of the stable simulation!)? Are the differences in the temperature profile close to the surface related to the way the PBL scheme is using the surface flux condition? Close to the surface MYNN seems more unstable (based on the temp. profile) than 3DPBL, yet the TKE is much lower for MYNN. Why are the results averaged over a 24-hour period (why not using 1 or 2 hours)?

In general, WRF simulations with different PBL schemes can produce different states of the atmosphere for a wide variety of reasons. However, in the case of pseduo-steady idealized simulations, it becomes much simpler to attribute the source of differences in atmospheric state. Here, the differences between MYNN and the 3DPBL primarily arise from different choices of closure constants and length scale formulations (L122). Recent studies using 3DPBL (Arthur et al. 2022; Juliano et al. 2022) suggest that model results under both stable and convective conditions are quite sensitive to these choices. Model sensitivity to these factors is actively being explored (Eghdami et al. (2022)). Thus, addressing each of these questions point-by-point:

- The difference in stable TKE profiles between MYNN and the 3DPBL likely arises because the 3DPBL and MYNN have different closure constants and length scale formulations.
- Hub-height TKE values are different between the two PBL schemes in stable and unstable conditions because we designed our experimental set up to match hub-height wind speeds instead of TKE values.
- As was the case in the stable simulations, the TKE profiles are different in the unstable simulations likely because of the different closure constants and length scale formulations.
- Temperature profiles are the same for MYNN and the 3DPBL in neutral and unstable conditions. They differ for stable conditions because the two PBL schemes were spun up for different durations, in order for them to reach matching hub-height wind speeds (Figure 1).
- As you have noted below in the minor comments, MYNN exhibits unexpected behavior in its unstable TKE profile---it unexpectedly produces more TKE in neutral conditions than in unstable conditions. The fact that stronger near-surface temperature gradients in MYNN lead to less TKE than in the 3DPBL is likely driven by this same confounding mechanism. On top of this, again the difference in closure constants and length scales will further drive the PBL schemes to predict different near-surface TKE levels.
- We chose to average over a 24 hour period because oscillations of hub-height wind speed lasted about 12 hours (Figure 1), and we wanted to smooth those out.

Minor comments:

1. Line 38 delete “more”

We have removed "more", and the line now went from "While this approach is more common for onshore sites" to "While this approach is more common for onshore sites".

2. Line 54 it should be “planetary”

Thank you for catching our typo. We have corrected it.

3. Line 57 are the traditional PBL schemes governing vertical turbulence fluxes or all turbulence fluxes (as it reads)
We have clarified the statement from "PBL schemes govern turbulent fluxes and mixing within the atmospheric boundary layer" to "PBL schemes govern turbulent fluxes (typically just vertical turbulent fluxes) and mixing within the atmospheric boundary layer" (L52-53).

4. Line 70 it should read “in order to”
Thank you for catching our typo. We have corrected it.

5. Lines 78-80 the sentence reads as NWP models use these WFPs, but is there a NWP model that uses Abkar’s?
Thank you for bringing this to our awareness. The WFP review paper (Fischerit et al. 2021) refers to the "Abkar WFP". However, we have not found an NWP paper that uses this WFP, so we have removed the reference to it.

6. Line 83 Replace “that some” by “that most if not all”
We have replaced "that some" with "that most if not all".

7. Lines 127-128 Why do the constants of the 3DPBL come from MY82 and not from the updated NN89?
The choice of closure constants is an active area of research for the 3DPBL. In this work, we selected the MY82 closure constants as they were tested in Juliano et al. (2022) whereas the NN89 constants were not tested.

8. Line 148. You need to explain why not using Archer’s suggestion for \alpha
We initially had an explanation for our choice in \alpha in the Methodology section for the real simulations, but the other reviewer told us to cut the explanation, ultimately because “Larsen and Fischereit (2021) themselves did not seem to trust the findings enough to make any recommendations about C_TKE based on their study”. Thus, following the other reviewer’s suggestion, we are not explaining our choice.

9. Table 1. There is a problem with the units of the heat flux
Thank you for catching our typo. We have corrected it.

10. Lines 170-176 All this wording can be omitted. I guess the real reason why you wanna match the hub height wind speed is to have the same Ct values
We have removed Lines 170-176 to make the paragraph more concise, thank you.

11. Line 180 Does it make sense to force sable simulations imposing surface fluxes? See Basu et al. (2008)
This is a very fair question. We chose to force the idealized stable simulations with a heat flux so that they could be compared to "typical" conditions in the offshore real simulations. This was

helpful for the comparison, as it is much stranger to try and calculate a "representative cooling rate" of the real simulations. In a context where real simulations are not considered, we agree that a temperature cooling rate would be preferred to a heat flux at the surface. However, reading Basu et al. (2008), we believe our simulations here are okay, because our simulations are only weakly stable. The conclusion of Basu et al. (2008) states "It is argued that any PBL model (single column or LES) will only be able to capture the near-neutral to weakly stable regime if surface sensible heat flux is prescribed… In order to represent the moderate to very stable regime in a boundary layer model, unquestionably one needs to use surface temperature as a boundary condition as shown in this paper." Because we are weakly stable, using heat flux should be acceptable.

12. Line 182 How do you estimate the height of the boundary layer?
We now clarify, "After spin up, the boundary-layer height as determined through the NWF temperature profile (Fig. 2) is approximately 250 m in the stable simulations, 550 m in the neutral simulations, and 600 m in the unstable simulations." (L172-173)

13. Line 242 well there is no real "turbulent hub height wind speed" in a PBL simulation
We have removed the word "turbulent" in order to be more precise.

14. Line 245 These spin ups will be much shorter if using lower top boundaries; also you will have perhaps less numerical noise and unwanted waves
While this may be true, we retain the tall upper boundary in order to be consistent with Fitch et al. (2012).

15. Lines 246-249 These lines can be omitted
Thank you for noting this. We elected to retain these lines, as we think the transition sentences are helpful for the reader so there is clear motivation on why we need to discuss NWF profiles before discussing wakes.

16. From the results in Fig. 2 it does not seem like the stable and unstable simulations stabilize
Indeed, the hub-height wind speed continues to change for the stable and unstable simulations, whereas it is near-constant for the neutral simulations. However, this non-stationarity is basically an unavoidable part of studying the ABL in stable and unstable conditions. For example, as long as we continue to apply a positive heat flux, the height of the capping inversion will continue to grow in the unstable ABL. Our averaged hub-height wind speeds are "pseudo-stationary" however, and as such, we believe they are sufficient to study averaged wake effects.

17. Line 266 Any idea why the TKE is weaker in the unstable than in the neutral MYNN? Could something be wrong?
This phenomena is something we have struggled with. MYNN is a thoroughly tested and well-established PBL scheme, so we were surprised to see this behavior as well. We are confident that our simulation inputs are correct (and they can be double checked through our open-sourced code).

18. Line 286 something is weid in “should be run that are”
Thank you for catching our strange wording. We have reworded "Thus, this variability within the idealized runs suggests that real-world case studies should be run that are tailored to a specific region and turbine configuration." to "Thus, this variability within the idealized runs suggests that real-world case studies should be tailored to a specific region and turbine configuration.".

19. Line 288 “external wakes” was introduced when you introduce the real cases but now you use it for idealized cases… which do not have external wakes… also line 296
Thank you for noting this inconsistency. We have updated the text to now use terminology like “the wake within the extent of the plant” and “the wake outside the extent of the plant”.

20. Figure 5 Is the side view a vertical cross section parallel to x? and where in y?
This specifically a side view of horizontally averaged wind speed deficits parallel to x. Deficits are averaged over the y-extent of the farm (L265).

21. Caption Fig. 6: it should read “…6a was calculated as the difference between the results in panels…” Also you say “are calculated in a similar manner”: this is not really true because you do not have tendencies in Fig. 5
We have updated the text from reading "Figure 5a was calculated as panel Figs. 4a-4g." to "Figure 5a was calculated as the difference between the results in panels Figs. 4a-4g." (caption for Figure 5). We have also updated the following sentence from "The u-tendency deficits in the 100TKE simulations are calculated in a similar manner." to "The u-tendency deficits in the 100TKE simulations are calculated using a similar procedure involving tendencies."

22. Line 163: this is interesting. Why Mangara et al. saw substantial changes in TKE so far downwind? There is no mechanism to preserve TKE so far I think
We are also surprised by the Mangra et al. results. Both our idealized and real simulations showed that TKE was advected a much shorter distance than what they saw. The real simulations of Siedersleben et al. (2020) also show TKE advection over a much shorter distance than Mangara et al. found.

23. Lines 364-366 But this observation is because in stable and unstable conditions the TKE in the NWF is nearly doubled in 3DPBL than MYNN?
Not exactly. While yes the 3DPBL produces more TKE than MYNN in stable and unstable conditions for the NWF scenario, the two PBL schemes produce near-identical TKE profiles in the neutral case. That is why L364-366 read “The 100TKE 3DPBL simulations also *consistently* predict stronger levels of additional TKE than their MYNN counterparts” (emphasis here only).

24. Figure 7: see my comment 20
To be clear in the consistency between figures, we have updated this caption to read "Same as Fig. 5, but for TKE..."

25. Figure 8: Why are the capacity factors so different for the turbines on the left of the array for neutral conditions compared to stable and unstable? They should be nearly the same at those turbines right?

This is a good point, thank you for catching it. These differences in the arise ultimately show up in the visualization because power production is extremely sensitive to velocity, following U^3. This is why power production even strictly within the leftmost column of any one simulation can substantially vary. For example, Fig 7a has capacity factor ranging between 54-66% in the upwind column. This sensitivity is then coupled with local variations in wind speed. The neutral wind speeds were fairly constant during the 24 hour performance period (Fig 1), whereas the stable and unstable winds were more variable. Because power production scales with U^3, the stronger variability in time results in an over-exaggeration of power production, even though all simulations have a 24-hour average wind speed near 9.35 m/s.

26. Lines 411-412 “The unstable… are 7.69 m/s… NWF profiles”. This is not needed.

We have entirely cut this part of the text, as the real simulations have been removed.

27. Figure 9: in neutral conditions, you most probably have the lowest friction velocities, which translates also in the lowest winds compared to the other stability cases. And this is kind of a problem as your stable case is probably biased by a large number of near-neutral cases; cases that would have fall into the category of neutral if the Obukhov length (instead of the flux only) was used.

As was the case for the previous comment, this section has been removed because the real simulations are gone.

28. Line 535: “MYNN predicted strong….. in stable conditions” This was not the case for the mid atlantic simulations, why? Maybe related to comment 27?

We note that stable MYNN simulations produced longer wakes than the 3DPBL in the ideal case (Fig 3 a,g) as well as in the real case (old Fig 11 a,g). However, the real simulations have been removed, so this point is moot.

References

Basu et al. (2008) An inconvenient “truth” about using sensible heat flux as a surface boundary condition in models under stable stratified regimes. Acta Geophysica, 56, 88—99

References

- Arthur, R. S., et al. “Improved Representation of Horizontal Variability and Turbulence in Mesoscale Simulations of an Extended Cold-Air Pool Event, Journal of Applied Meteorology and Climatology” 2022, 61(6), 685-707.
- Eghdami, Masih, et al. “Diagnosis of Second-Order Turbulent Properties of the Surface Layer for Three-Dimensional Flow Based on the Mellor–Yamada Model.” Monthly Weather Review, vol. 150, no. 5, May 2022, pp. 1003–21. journals.ametsoc.org, https://doi.org/10.1175/MWR-D-21-0101.1.

- Fitch, Anna C., et al. "Local and Mesoscale Impacts of Wind Farms as Parameterized in a Mesoscale NWP Model." Monthly Weather Review, vol. 140, no. 9, Sept. 2012, pp. 3017–38. journals.ametsoc.org, https://doi.org/10.1175/MWR-D-11-00352.1
- Fischereit, Jana, et al. "Review of Mesoscale Wind-Farm Parametrizations and Their Applications." Boundary-Layer Meteorology, Aug. 2021. Springer Link, https://doi.org/10.1007/s10546-021-00652-y.
- Juliano, Timothy W., et al. "'Gray Zone' Simulations Using a Three-Dimensional Planetary Boundary Layer Parameterization in the Weather Research and Forecasting Model." Monthly Weather Review, vol. 1, no. aop, Oct. 2022. journals.ametsoc.org, https://doi.org/10.1175/MWR-D-21-0164.1.
- Siedersleben, Simon K., et al. "Turbulent Kinetic Energy over Large Offshore Wind Farms Observed and Simulated by the Mesoscale Model WRF (3.8.1)." Geoscientific Model Development, vol. 13, no. 1, Jan. 2020, pp. 249–68. gmd.copernicus.org, https://doi.org/10.5194/gmd-13-249-2020.

Second review of "The Sensitivity of the Fitch Wind Farm Parameterization to a Three Dimensional Planetary Boundary Layer Scheme" by Rybchuk et al., submitted to Wind Energy Science Discussions

The authors have made concrete efforts to improve the paper and address the reviewers' concerns, including my own. I am especially glad that the authors checked the TKE advection issue and that they showed strong evidence that it was indeed active in their runs. That was my major reason of concern and it has been satisfactorily addressed. As for the extremely long reply about the revisited motivation of the study to address uncertainty quantification (as opposed to validation), I could counter-argue that what the authors did is not exactly what Archer et al. (2014) recommended, which was basically to use ensembles. They meant several ensemble members, like 15-20, but here there are only 2 (MYNN and 3DPBL schemes). A 2-member ensemble is not sufficient to characterize uncertainty, but I will grant that the authors found a substantial difference in power, which is interesting and good to know, thus I will let this go.

Dear reviewer, thank you for your continued thoughtful feedback on our manuscript. We have updated the manuscript in accordance with your suggestions. We note that following feedback from the editor and Reviewer 1 in this latest round of revisions, we have cut nearly everything regarding the real simulations. We will also note that as the review process has been going on, we have found and fixed the bug regarding strange near-surface TKE values in the idealized 3DPBL simulations. Thankfully, as we anticipated and noted earlier in the review, this bug had negligible effects on wind speed and TKE profiles.

Replies to previous comments

3. 101: I think I know what you are trying to say, but it needs to be defined better because an external wake cannot be defined as a "distance". Also, here you use 0.2 m/s as the threshold, but in the rest of the paper it seems to be 0.5 m/s (e.g., Figure 3 and 11, dashed blue line).

Following feedback from Reviewer 1, the literature review section has been removed. While the text has been removed, we want to address the challenge of characterizing external wakes here, as this feedback is also brought up later. As the recent WFP literature review paper by Fischereit et al. (2021) states
"One challenge that we identified is that from our review there is no standardized or common definition of a recovery length behind a farm. Studies used for instance the efolding distance (Fitch et al. 2012), the location of 2% difference between a simulation with and without wind farms (Pryor et al. 2020) or the location where the wind speed has recovered to 95% of the freestream wind speed (Cañadillas et al. 2020). Due to this variety of different definitions, it is difficult to compare wake lengths across studies quantitatively."
Each of these studies selects a definition of an external wake that is reasonable for the question they are studying. In our analysis, we study wakes from large farms of large turbines. As such, we select three metrics: 1 m/s threshold, 0.5 m/s threshold, and the efolding distance because we expect large wakes. This is imperfect, but we also acknowledge this challenge that the field faces.

I think that the authors misunderstood my comment. I did not criticize the use of a threshold at all, that was and is just fine. I just found it confusing that the authors used two different values (0.2 m/s and 0.5 m/s) in two parts of the same manuscript. I actually wonder if 0.2 was a typo perhaps? In fact, you no longer use 0.2 m/s in the revised manuscript. Not important anyway, but I thought I would clarify what I meant.

Thank you for clarifying.

4. Table 1: the same label here is used to indicate three different runs. Please use unique labels for each run, like "S-NWF" for stable, "U-NWF" for unstable etc.

While it is entirely reasonable to label the idealized runs "S-NWF" and "U-NWF", it would not be reasonable to name the monthlong real run in a similar manner. Thus, for the sake of consistency between the ideal runs and the real runs, we retain the original label format.

I am sorry, but I disagree. You must not use the same label to indicate different runs, period. The point of a label is that it identifies uniquely the run you are talking about. However, the labels are not even used in the manuscript, so this discussion is moot. I would just suggest that the column "Label" be actually renamed "PBL scheme" and the prefix "NWF" be removed because this table is already for no-farm simulations.

Thank you for the feedback. We have updated Table 1 so that the column "Label" is now called "PBL Scheme" and we have removed the "NWF" prefix.

8. 322: Why 0.5 m/s deficit if 0.2 m/s was stated earlier?

As stated in response to Minor Comment 3, there is no standard for external wake characterization and 0.5 m/s threshold makes sense for the scale of problem we are studying.

Again, I was not criticizing the value, but noting the contradiction of using two different values in two different parts of the manuscript. All resolved now since there is no longer a 0.2 m/s threshold.

15. 476: Are these results with 0% TKE or 100% TKE? Why not 25% TKE as recommended?

These results are with 100% TKE. Thank you for noting that omission, and we have updated the Methods section to include that detail. "All wind plant simulations are run with α=1. While validation of this parameter is limited, we note that Larsen and Fischereit (2021) saw more accurate results in an offshore wake study with that value (α=1) than the value of α=0.25 recommended by Archer et al. (2020)." (L232-234)

Actually, I read Larsen and Fischereit (2021) very carefully and they did not reach any such conclusion in their paper. They even explicitly stated that "It remains inconclusive … what are the correct C_TKE values to use; more measurements are needed for further investigation." The authors might be referring to Figure 14 in Larsen and Fischereit (2021), which shows very large TKE injection over the wind farm during one 2-hour flight, so large that none of the parameterizations, no matter which settings were used, could capture it. In particular, the fact that not even the results with advection turned off, which are well known to cause an overshoot of TKE in the grid cells of the wind farm because TKE has no way to go but continue accruing, could match the observed values indicate that perhaps something was off with the methods used to calculate TKE from the measurements. Even Larsen and Fischereit (2021) themselves did not seem to trust the findings enough to make any recommendations about C_TKE based on their study. Anyway, I do not intend to conduct a review of Larsen and Fischereit (2021) here, but I think I provided enough evidence to suggest that the sentence at lines 232-234 should be removed.

Following feedback from everyone involved in the peer-review process, we have cut any mention of the real simulations, and as such, L232-234 have been removed.

17. Figure 12: as in #14, not OK to have 4 colorbars.
Figures with multiple colorbars are employed within academic literature. For example, see Figure 3 in Pryor et al. (2020) and Figure 7 in Brugger et al. (2022).

While other papers may have had good reasons to make such a choice, I really do not see the advantage of 4 colorbars here. It is extremely hard to compare results from one heatmap to the next. The only advantage is that one could discern patterns within the wind farm, but the authors do not actually discuss any such patterns in the text, so there is really no advantage.

As was the case for the above comment, nearly all mentions of real simulations have been cut from the manuscript, and as such, Figure 12 and its 4 colorbars have been removed.

New comments

1. P. 3, l. 79: the paraterization by Abkar and Porte-Agel does not have a name and calling it the “Abkar WFP” is presumptuous. The parameterization by Pan and Archer (2018) also is not called the Pan WFP, but it has a name, it is called the “hybrid” WFP.

We called it the Abkar WFP following Fischeit et al. (2021), but we have removed mention of this WFP following feedback form the other reviewer. We have now also updated "Pan WFP" to the "hybrid WFP".

2. P. 3, l. 89: your research question has changed to “How sensitive are modeled mesoscale wakes to the choice of PBL parameterization?”, which is a much better fit to the study, but I think you should be a bit more humble and admit that you have only examined one additional PBL scheme, one that has not really been validated much because it has not even been released officially with WRF as far as I know. So I would recommend that you add a few sentences to explain your choices along the lines of: “Ideally, we should test all 13 PBL schemes with the Fitch WFP to fully characterize the uncertainty. Here we propose, as a first step, to compare two PBL schemes: the conventional NYMM and the newly-developed NCAR 3DPBL. We chose the latter because …. “ and explain why you chose it over the other 13 schemes? A few sentences suffice, like the fact that it has a prognostic equation for TKE.

We wholeheartedly agree with your framing and we have updated the closing paragraph of the introduction accordingly (edits in italics): "In this paper, *we begin to* address the question: How sensitive are modeled mesoscale wakes to the choice of PBL parameterization? *Ideally, this question would be addressed by studying all 13 PBL schemes in WRF with the Fitch WFP insofar as that is possible. Here, as a first step, we compare two PBL schemes*: MYNN (Nakanishi and Niino 2009) and the recently developed NCAR 3DPBL (Kosovic et al. 2020, Juliano et al 2022). *We chose the latter as it as a prognostic equation for TKE, which is important as the Fitch WFP modifies TKE fields.*"

3. P. 6, l. 168: good idea to match the hub height wind speed rather than the geostrophic speed.

Thank you.

4. P. 9, l. 233-235: as explained above, this sentence should be removed because it is not consistent with the recommendations and findings reported by Larsen and Fischereit (2021) themselves.

This sentence has been removed.

5. What is “pp”? It is used only in the last sections of the paper (that is why I did not notice I earlier) but is not defined. Perhaps it means just “%”, which was used in the abstract and other sections? Please use one convention only.

“pp” is percentage points (defined in L224). It is used in early parts of the paper to take the difference in wake wind speed deficit percentages (e.g. 14% - 12% is 2pp) as well as later to discuss differences in capacity factor, which is also expressed as a percentage.