Reviewer 2

The paper inserts for the first time the Fitch wind farm parameterization in the newly developed 3DPBL scheme in the WRF model. This is important and innovative because, as far as I know, the Fitch parameterization has only been coupled with the MYNN PBL scheme. It is therefore valuable to see how it would work with a different PBL scheme. However, the paper consists of pages and pages and pages of detailed and rather pointless differences between the results obtained with the two schemes, first with idealized cases, then with a series of simulations of a few offshore wind energy areas in the US Northeast, leaving however no useful information on which would be better for which cases and why. I am afraid that, in order for the paper to be acceptable in its current format, too much additional work would be required (i.e., redo all idealized runs and simulate a different real farm), as discussed next.

An alternative would be to remove Section 4 entirely. Adding a real case would be valuable if it allowed the authors to validate the 3DPBL+Fitch coupling, but it has no value in this manuscript unfortunately, it just adds pages and pages of minutia and repetition.

Dear reviewer, thank you for taking the time to read the manuscript and provide feedback. We have updated the manuscript following the feedback from both reviewers, and we believe the new manuscript is significantly stronger. We have substantially cut down on the amount of “detailed and rather pointless differences”. Following feedback from Reviewer 1, we have rerun and significantly revised the idealized simulations to (1) match hub-height wind speeds and (2) use a roughness length for offshore conditions. Please find our comments below in red.

Major points

1. Although some of the co-authors have access and/or have participated in field campaigns that have collected plenty of data on wind farm wakes, inland and offshore, and on observed wind farm power production (e.g., Siedersleben et al. 2020 just to mention one), no comparison against any type of observations is offered in this study. Why did the authors choose to simulate the Vineyard Wind and the other U.S. wind energy areas, for which no data are available yet, when so many other farms with data are available? At a very minimum, high-resolution simulations (like HRRR) could have been used for the wind speed profiles for August 2020 for Figure 9. But, better yet, a different farm with actual wake observations should have been simulated instead.

1. The motivation for this manuscript could have been clearer and stronger.

Both reviewers noted that this manuscript was not engaging, so we have updated the introduction to more clearly highlight the important and time-sensitive problem that our research addresses (L25-49). Our primary motivation for conducting this analysis is as follows: offshore wind resource assessments are necessary for the rapidly developing offshore wind industry, but these resource assessments suffer from a lack of quality observations across most of the U.S. Thus, the offshore research agenda within the U.S. has explicitly solicited researchers to improve uncertainty quantification for offshore wind resource assessments. This call has come from academia (e.g., Archer et al. (2014), "Meteorology for Coastal/Offshore Wind Energy in the United States Recommendations and Research Needs for the Next 10 Years", which states that Research Need #2 is Uncertainty Quantification, especially in the form of ensemble simulations, which is what our research enables) as well as U.S. federal scientific agencies (e.g., Shaw et al. (2019) "Workshop of research needs for offshore wind resource characterization", in which the need for uncertainty quantification is stressed time and time again). Our research directly addresses the need to improve uncertainty quantification of offshore wind resource.

2. Our article improves the capability for uncertainty quantification, which is distinct from model validation, but also crucial to ensure reliable numerical models.

As addressed above, there is an established need to quantify uncertainty in numerical models of wind resource, especially for regions that lack high quality observations. The National Research Council of the National Academies put out a book that describes the importance of (and distinctions between) uncertainty quantification and validation, "Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification". They state that validation answers the question of "How accurately does the model represent reality for the quantities of interest?" and uncertainty quantification addresses "How do the various sources of error and uncertainty feed into the uncertainty in the model-based predictions of the quantities of interest?". These are interrelated but distinct and co-equal parts of the verification, validation, and uncertainty quantification (VVUQ) process. The book states that "the first UQ task is to quantify uncertainties in model inputs", and we do precisely this by developing new WRF code that allows users to vary the PBL scheme in a wind farm simulation. In our paper, we demonstrate that indeed there is a great deal of wind resource uncertainty that is associated with the choice in PBL scheme. Perhaps our most simple and most important finding is that "the 3DPBL generates 4.7%-7.8% less power than MYNN in August 2020". This finding could have serious implications for the financial viability of the offshore wind industry, and we demonstrate this finding without employing observations.

We concede, as the book notes, that "VVUQ tasks are interrelated". However, our work (1) enhances the ability of researchers to conduct VVUQ studies for wind farm simulations in the future and (2) demonstrates for the first time that, indeed, future VVUQ studies should vary the PBL scheme. Both reviewers note that our paper already feels too long, and as such, it would not be possible to additionally conduct a thorough validation study that meets the best practices delineated in the book, such as

- “Principle: Validation assessments must take into account the uncertainties and errors in physical observations (measured data).”
- “Best practice: If possible, use a broad range of physical observation sources so that the accuracy of a model can be checked under different conditions and at multiple levels of integration.”
- “Principle: Validation and prediction often involve specifying or calibrating model parameters.”
- “Principle: The uncertainty in the prediction of a physical [quantity of interest] must be aggregated from uncertainties and errors introduced by many sources, including discrepancies in the mathematical model, numerical and code errors in the computational model, and uncertainties in model inputs and parameters.”

We believe that a future article that carries out such a validation study and meets these principles would indeed be valuable, especially when conducted in conjunction with observations from upcoming field campaigns that include focused efforts to characterize mesoscale wake effects (e.g. AWAKEN/ARISE).

<u>3. Now that our idealized simulations do not share large-scale forcing, it is increasingly important that we simulate a real case study with both MYNN and the 3DPBL that share large-scale forcing.</u>

As discussed in greater detail below, following feedback from Reviewer 1 we have re-run all the idealized simulations so that their NWF hub-height wind speeds match. This necessitated tuning the large-scale forcing for each simulation. As such, the idealized simulations and real simulations highlight two distinct effects now. The idealized simulations explore the question of “How does the unique momentum recovery parameterization of each PBL scheme affect wakes”, and (as we discuss in greater detail in the manuscript) the real simulations ask “How do differing predictions of hub-height wind speed affect wakes?”. Thus, the real simulations that we have run take on even greater importance now.

<u>4. The scope of our article is consistent with the scope of other published WFP sensitivity studies, some of which have been published in *Wind Energy Science*.</u>

- Bodini et al. (2021) published a WRF wind resource assessment study in *WES* (no WFP, but ultimately addressing the same fundamental question of uncertainty associated with a PBL) in the offshore US that did not compare to observations. This paper employed a 16-member WRF ensemble that was generated using in-built WRF capabilities.
- Pryor et al. (2021) published a WRF WFP sensitivity in onshore US in *JAMC* that did not compare to observations.

Thus, our methodology is consistent with other academic publications that address the same fundamental question we investigate.

5. Itemized response to the concerns raised here

Thus, in summary:

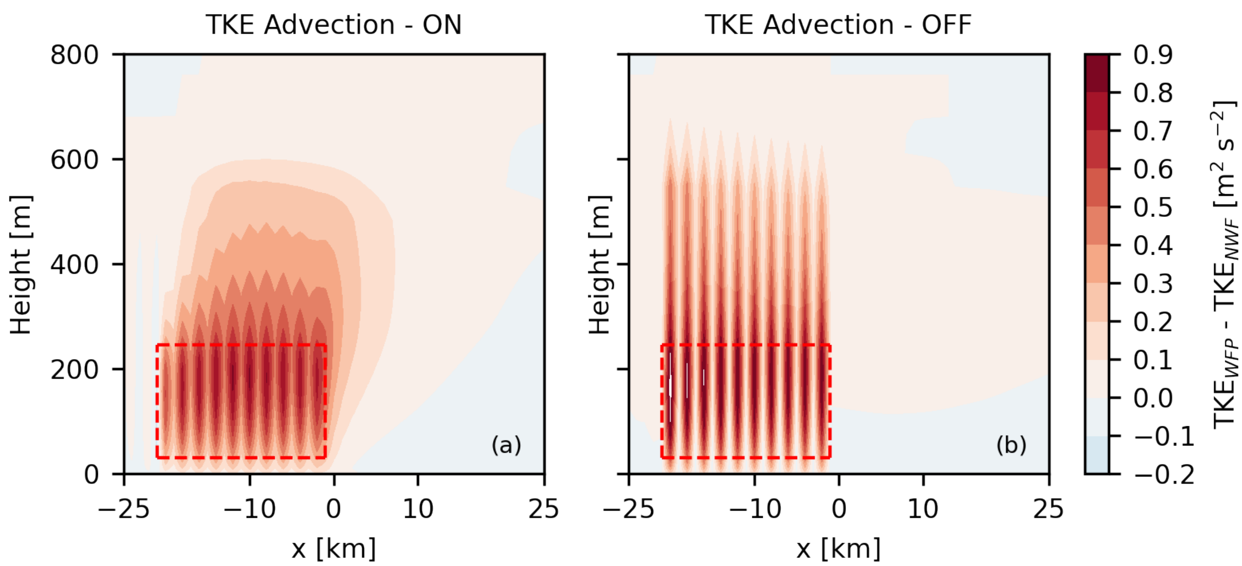
- " Why did the authors choose to simulate the Vineyard Wind and the other U.S. wind energy areas, for which no data are available yet, when so many other farms with data are available? ": We conduct an uncertainty quantification study in this paper. Our work (1) enhances the ability of researchers to conduct VVUQ studies for wind farm simulations in the future and (2) demonstrates for the first time that, indeed, future VVUQ studies should vary the PBL scheme. We believe future validation studies will be important and complement the uncertainty quantification analysis undertaken here.

2. The authors state that TKE advection is turned on (see l. 243), but it does not seem to be true. Figure 6 shows without doubt that all the added TKE is confined within the boundaries of the wind farm and above it, but not advected downwind at all. With the MYNN scheme in particular (top two rows), one can even see the individual positions of the turbines, one every other grid cell, with the added TKE at their grid cells and above, but none in the next adjacent cells downwind. This proves that no advection is actually operating. The authors need to double check that bl_mynn_tkeadvect is indeed set to true in the inner domain. Since TKE advection appears to be wrongfully turned off in all the simulations, all the conclusions of the paper are potentially invalid.

We thank you for your attention to detail regarding our methodology. We have verified that bl_mynn_tkeadvect was indeed turned on and functional:

- Looking at the idealized, neutral 100TKE namelist for example (https://github.com/rybchuk/wfp-3dpblsensitivity/blob/main/runs/idealized/neutral/mynn-tke/namelist.input), we verify that "bl_mynn_tkeadvect = .true., .true.,"

- We have verified that the WRF code modifications from Issue 1235 are in the version of WRF that we use
- While TKE quickly recovers downwind of the turbines at the rotor disk heights, TKE is advected downwind above the rotor disk, most clearly seen in Figure 6a,b,d,e
- The prognostic version of the NCAR 3DPBL (which we use in all our simulations) inherently advects TKE (independently of bl_mynn_tkeadvect). The fact that the MYNN and the 3DPBL produce similar TKE addition plots futher suppors that MYNN has TKE advection turned on
- Finally, to definitively illustrate that TKE advection was turned on in the simulations in the manuscript, we have run a supplemental MYNN idealized neutral simulation in which the TKE advection was turned off. Figure RF1a has TKE advection turned on, and TKE is visibly advected downwind. Figure RF1b has TKE advection turned off, and TKE remains within vertical columns.



Response Figure 1. A TKE side view of idealized neutral simulations for which TKE advection is turned on (a) and turned off (b).

Thus, we are confident that our simulations in the manuscript were run with TKE advection turned on.

Minor points

1. 55: There is another wind farm parameterization for WRF in the literature: the hybrid model by Pan and Archer (2018).

We have updated the text to include a citation to Pan and Archer (2018) (L79).

2. 64 and 133 and 164: Missing citation "?"

Thank you. We have corrected the missing citations.

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 suse 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 e-folding 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 e-folding distance because we expect large wakes. This is imperfect, but we also acknowledge this challenge that the field faces.

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.

5. 203: not OK to cite a manuscript in preparation, please remove Rosencrans et al.

We have removed the reference to Rosencrans et al.

6. 206: type for "pseudo"

We have corrected the typo, thank you for catching it.

7. 209: How many turbines are there in total? 25 perhaps?

There are 100 turbines in the idealized simulations, as denoted by "The second case (100TKE) includes a 10-by-10 grid turbines based on the of 12-MW International Energy Agency" (L186).

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.

9. 322: I cannot understand what the e-folding distance is. Please include an equation. To be honest, I do not even understand why this variable is even introduced, it does not add much, it is overly sensitive to the stability and choice of the scheme, and it is no longer used in the real simulations later. Consider dropping it since it does not add much.

The e-folding distance is a relative measure used to characterize external wakes. The metric was introduced in the original Wind Farm Parameterization paper (Fitch et al., 2012), and as our idealized simulations parallel much of the analysis in that paper, we feel that it is appropriate to include the metric. We provide a formula on how to calculate the metric: "we calculate the e-folding contour as 1/e times the average internal wake strength, or about 36% (Fitch et al., 2012)" (L289)

10. Figure 3: I am surprised that the maximum deficit possible is 1 m/s (note that the maximum deficit is 4 m/s in Figure 11). This must be the most efficient ideal wind farm ever designed. Why is the flow from the west-southwest? I would recommend using white for the range -0.25 – 0.25 m/s.

- The new, idealized offshore simulations show maximum deficits in excess of 2 m/s. The mid-Atlantic simulations show stronger deficits because (a) there are substantially more turbines (1418 in LEASE, 177 in VW-ONLY, and 100 in the idealized simulations) and (b) Vineyard Wind is longer with respect to the dominant wind direction, thus leading to stronger internal waking.
- The flow in the idealized simulations is from the west-southwest because of the combination of friction and the Coriolis effect. The same effect was observed in Fitch et al. (2012). We have updated the paper to mention this effect: "Wakes are rotated from the U-geostrophic wind due to the combination of friction and the Coriolis force." (Figure 4 caption).
- We retain the original colorbar values for the range –0.25 - 0.25 m s-1 so that the subtle but real acceleration in the stable idealized simulations is visualized.

11. 345-350: I find it **very** difficult to believe that the addition of TKE causes a longer wake. Also very confusing that the weird decrease in TKE in one specific case (Figure 6g) can be used to explain this general and counter-intuitive finding. To me this is another flag that suggests that advection of TKE was **not** turned on.

Please see our response to Major Comment 2 for analysis that demonstrates that TKE advection was turned on.

12. 384: not OK to cite a manuscript under review. Please remove Bodini et al.

Bodini et al. has been published during the review of this manuscript, and the reference has been updated accordingly.

13. 409: the authors themselves note that there is no advection of TKE! This is not a realistic result. Flag bl_mynn_tkeadvect must be true for TKE to be advected, at least with the MYNN scheme.

We state that "tend[s] to not advect *far* downwind" (emphasis mine). Please see our response to Major Comment 2 for analysis that shows that TKE advection was turned on.

14. Figure 8: please use one color scheme! You can intervals that are variable to better emphasize features, but using two colorbars like that is not OK.

The updated idealized simulations all have similar hub-height wind speeds, and correspondingly, similar power production. Figure 8 now only uses one color scheme.

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)

16. 484: define "centroid"

We have updated the text to read "we calculate average profiles at the middle ofVineyard Wind 1."

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).

18. 28: by this point, I could not force myself to read the manuscript anymore. Too boring and pointless. This section on the real cases is rather useless without observations and does not add anything to the discussion of the idealized cases. The paper would be better off without Section 4.

We hope the reviewer will be willing to complete its review now that we have clarified some points (e.g., that TKE advection was indeed turned on) above and that the manuscript has been improved. As we would like to stress again, our U.S. east coast analysis reveals important information that a number of stakeholders (scientists, government planners, industry) care about. Perhaps the most important and the most simple finding of the real analysis can be found in our abstract: “the 3DPBL generates 4.7%-7.8% less power than MYNN in August 2020, depending on the turbine build-out scenario”. Our study shows that the choice of PBL scheme could lead to significantly different AEP predictions, and in theory, the choice of wind farm could flip AEP from being in a profitable scenario to an unprofitable scenario.