Response to Referee #1

Referee comments appear in black and author responses appear in blue.

This is a WRF modeling study of the wakes of future offshore wind energy areas planned along the US east coast. It considers three scenarios of offshore wind development: Vineyard Wind only, lease areas in the Mid-Atlantic, and lease + call areas in the Mid-Atlantic. An impressive modeling effort was undertaken, with nested high-resolution WRF runs for one year repeated for the three scenarios above, plus the control case with no farms, plus runs with 0% and 100% TKE added. The team had also numerical noise issues, as seems to be the norm when the wind farm parameterization is turned on, and therefore had to deal with several additional runs to take care of it. The paper is definitely worth publishing as it includes interesting and valuable results. It is exceptionally well written. It was a pleasure to read such a good paper!

We thank the reviewer for reading thoroughly and providing thoughtful suggestions to improve this article.

Major Comments

1. **The study focuses too much on added TKE.**

The paper promises to study annual variability of wake impacts (see comment #4 below about how it is monthly and diurnal, not annual).

Thank you for pointing out our inaccurate use of word choice. Annual variability refers to multi-year studies while our research focuses on one year. You leave more title-specific suggestions, including this one, in comment #4, so we refer to all title changes in our response to comment #4.

Such a study should have a main run with certain fixed parameters, perhaps a few case studies of special interest or a validation effort, and a few case studies to assess the sensitivity to some of the parameters. Instead, in this study a lot of effort was put on the sensitivity.
The parameter of focus is the amount of added TKE, which is 25% in the default settings of the WRF model, and which was found here to have a relatively small impact on the power output (<5%). One would expect that the main run would be with 25% TKE and then a few cases (perhaps one week in each season) would be run with 0% and 50% and 100% TKE, in addition to 25% TKE, to assess sensitivity. The main results reported in the abstract and in the conclusions would be obtained with the default 25% TKE and a sentence or two would address sensitivity to TKE.

Thank you for the suggestion. In planning our simulations, we had extensive discussions about the merits of 0% vs 25% TKE, and finally decided to use 0% as a bottom limit rather than 25%. The 25% recommendation was based on only one study. Archer et al. (2020) recommended the use of 25% TKE based on idealized conditions, with neutral stratification, and for a one-wind-turbine setup. The 0% added TKE is more similar to the impact in the Volker et al. parameterization which has been used in several studies.

Instead, on one hand the study focuses excessively on the sensitivity to added TKE, because all the runs have been repeated entirely for 0% and 100% TKE, when a few weeks would have been sufficient. Of the 6 figures in the paper that describe modeling results (Figures 8-13), all of them are doubled to show 0% and 100% TKE. This would be understandable if TKE had a large impact on power output, but it did not (at most 5%).

Given the extensive discussion of the value of added TKE in the literature, we wanted to thoroughly document its variability. And, as pointed out above, 0% is more similar to Volker et al. and the 0-100% range therefore includes not just the Archer et al. suggestion but Volker et al. as well.

On the other hand, the team did not perform a year run with the recommended value of 25% TKE.

As noted above, the recommendation of 25% comes from one study, and we sought to provide more extensive bounds on the variability that could be introduced with a range of added TKE values. Archer et al. (2020) recommended the use of 25% TKE, and while extremely helpful to pioneer a suggestion for this issue, that recommendation is unfortunately limited in application for being run under idealized conditions, with neutral stratification, and for a one-wind-turbine setup. For this reason, Archer et al. reported that 25% TKE was the best choice for their setup, and further investigation is still required. There is uncertainty on what the "rule of thumb" TKE amount should be in regional wind plant modeling, and our results, because we explore the whole range of possibilities from 0% to 100%, provides a useful contribution by quantifying the (small) size of the impact of the TKE term.

A few weeks of simulation time may have been sufficient for future model development choices. However, the goal of our report was to provide the first year-long assessment of wake effects on power production, which is a highly sought-after dataset for industry partners and stakeholders.

Thus in principle every value that they report in the abstract should be a range, but it is not. Plus, no TKE results are shown (I would like to see the equivalent of Figure 8 but for added TKE). In summary, this study focuses excessively and at the same not enough on TKE. An obvious recommendation would be to ask the team to conduct a new one-year run with 25% TKE and rewrite the paper to focus on those results and reduce the sensitivity analysis. But I think that this would be an excessive request, plus there is already a lot of value in the current runs.

We appreciate that the reviewer recognizes that another set of 25% TKE simulations is computationally infeasible. Due to computational limitations, we cannot run an additional year-long simulation.

My first recommendation is therefore that the results in the abstract and conclusions, which do not report a range (i.e., the range of results with 0% and 100% TKE) but are presented without explanation as one value (see comment below), be modified by either reporting always the range, or by using an interpolation based on the few days of 25% TKE results that the authors have already run (Fig. A1).

We have modified the abstract as follows to incorporate the range of values:

> "Using a series of simulations with no wind plants, one wind plant, and complete build-out of lease areas, we calculate wake effects and distinguish the effect of wakes generated internally within one plant from those generated externally between plants. **We also vary the amount of added turbulence kinetic energy (TKE) between 0% and 100% to provide some uncertainty quantification.** The strongest wakes, propagating 55 km, occur in summertime stable stratification, just when New England's grid demand peaks in summer. The seasonal variability of wakes in this offshore region is much stronger than diurnal variability of wakes. Overall, year-long wake impacts reduce power output **by a range between 38.2% and 34.1% (for 0%-100% added TKE)**. Internal wakes cause greater year-long power losses, **from 29.2% to 25.7%,** compared to external wakes, **from 14.7% to 13.4%**. The overall impact is different from the linear sum of internal wakes and external wakes due to non-linear processes. Additional simulations quantify wake uncertainty by modifying the added amount of turbulent kinetic energy from wind turbines, introducing power output variability of 3.8%. Finally, we compare annual energy production to New England grid demand and find that the lease areas can supply **58.8% to 61.2%** of annual load."

Further, the conclusions are modified similarly:

- We now report "The average yearlong power deficits at Vineyard Wind considering internal wakes and external wakes from the LA range between 38.2% (TKE_0) and 34.1% (TKE_100)."
- Text is rewritten to include the range by "Yearly averaged wake losses induce power deficits at Vineyard Wind from 38.2% (TKE_0) to 34.1% (TKE_100)".

What I mean is that the team could obtain a relationship between average power output (or wind speed deficit or whatever the parameter of interest is) with 0%, 25%, and 100% TKE from the few simulated days. This relationship does not seem to be linear from Fig. A1. An example of this relationship might be something like: the power at 25% TKE is the mean of that at 0% and 100% TKE, on average. Then use that relationship to report one value (per parameter) in the abstract and conclusions, that "fitted" to 25% TKE.

As the reviewer has pointed out, the relationship between parameters and the amount of added TKE is a nonlinear relationship and so we have chosen to provide the range of values as above.

The second recommendation is that the authors add a figure and discussion on the TKE distribution in the wakes with 0% and 100% TKE, like Fig. 8.

We have added a section for the results and discussion of TKE at the hub height in new Appendix E, similar to Figure 8 (new Figure 11).

2. **The calculation of the losses from external wakes may be incorrect**

From the abstract, the effects of internal wakes are reported to be -27.4% and the effects of external wakes are -14.1%. The sum of the two is -41.5%. However, the combined effect is reported to be -35.9%. This is problematic. At first sight, this discrepancy may be the result of the non-linearity of the wake processes. If so, all the authors need to change is to reverse the order of two sentences and add a few words in the abstract to explain it: "Internal wakes alone cause greater year-long power losses (27.1%) compared to external wakes (14.1%). When both are present, however, the mean year-long wake impacts reduce power output by 35.9%, which is lower than the sum of the two due to non-linear processes."

Yes, we have noticed and discussed this nonlinear behavior, and have added a sentence to the abstract to explicitly note this behavior.

However, I suspect that there might be a design issue in the way the power losses are calculated in Eq. (9) and (10). Aside from the unclear notation (see comment below), the denominator of the two equations is not the same and that may be why the discrepancy arises. Eq. (10) is correct because there is no double counting: there are no losses in the denominator and the internal losses are only in the numerator. In Eq. (9), however, there are internal losses in both the numerator and the denominator, and they are not equal. The internal losses are not equal in the VW and CA cases because, as upstream conditions change due to external wakes in the CA case, the internal wakes change too and therefore the internal wake losses do not "cancel out", there is still some influence from the internal wake losses. As such, the ratio in Eq. (9) does not quantify just external losses because it still contains the effect of internal losses; it quantifies a mix of internal and external losses.
I suggest that the authors report Eq. 10 first (Loss_internal). Then, they should replace P_VW at the denominator with P_NWF in Eq. 9, to obtain the total effect from internal and external wakes due to the CA areas (call it Loss_total). We know this value: it should be -35.9% (from l. 384). The effect of the external wakes then is the difference between the total losses and the value from Loss_internal:
Loss_external = Loss_total – Loss_internal = -35.9% - (-27.4%) = -8.5% (Eq. 11)
This way the denominator is the same and the individual values for external and internal sum up to the correct total.

Thank you for this suggestion. There are several different methods for calculating the wake impact, and we have supplemented an additional method for calculating external losses as the difference between the total and internal losses, via a new equation (11).

Power losses from external, internal, and the total wake effects are calculated from:

$$Loss_{external} = 100 - \left(\frac{P_{LA,CA}}{P_{VW}}\right) * 100\% \quad (9)$$

$$Loss_{internal} = 100 - \left(\frac{P_{VW}}{P_{NWF}}\right) * 100\% \quad (10)$$

$$LOSS_{total} = 100 - \left(\frac{P_{LA,CA}}{P_{NWF}}\right) * 100\%\,, \quad (11)$$

where $P_{LA,CA}$ is the power production at Vineyard Wind grid cells in the presence of wakes by either the LA or the CA, P_{VW} is the power production in the presence of internal wakes from VW, and P_{NWF} is the power production from coupling hub-height wind speeds to the power curve. These methods are performed separately by added TKE amount. **We note that the upwind conditions change in a LA or CA scenario, due to external wakes, which can modify the internal losses in the numerator of Eq. 9. Thus, we provide an alternative method for calculating the external power losses as the difference between the total losses and the internal losses:**

$$\boldsymbol{LOSS_{external} = LOSS_{total} - LOSS_{internal}} \quad (12)$$

3. **The stability classification is not adequate.**

The authors use a very simple classification for stability based on the value of L (Eq. 8). Neutral conditions are those with abs(L)>1000 m. This is inconsistent with the published literature, e.g., Gryning et al. (2007) and Sathe et al. (2011) used 500 m, Wharton and Lundquist (2012) used 600 m, Rajewski et al. (2013) used 400 m, Archer et al. (2016) used 500 m. In fact, too few neutral cases were found here, less than 2.5% of the time (p. 29 l. 595). I am unsure what to recommend here because there is not an "accepted" value of L for neutral conditions, but the authors need to assess the sensitivity of their results to a few values, at a minimum 500 m. This could possibly help with the previous inconsistencies in the areal extent and wake length, as days that were actually neutrally stratified may have been mixed in with days with other stabilities to obfuscate some of the relationships.

As the reviewer acknowledges, there is a wide range of thresholds that have been used to determine stability regimes and there is not an accepted value of L to demarcate the line between neutral and stable or unstable conditions. The threshold of 1000 m is consistent with the published literature as Muñoz-Esparza et al. (2012) use this cutoff for neutral conditions in the offshore environment. (Most of the references cited by the reviewer were for onshore conditions). Our finding that neutral stratification occurs 2.5% of the time is only for the CA simulations, which is a subset of August-September of 2019 and June-July of 2020. For the yearlong period, our original reported number is double this value, at 4.48% of the time.

Additionally, through discussion with other WRF modelers, we learned that the WRF-output Obukhov Length (which we were using in the original calculations) is not accurate because it is calculated in the timestep before the heat flux is calculated. We have recalculated the Obukhov length directly using model-output variables at the same location (all figures and calculations incorporating stratification have been updated). The new percentages of occurrence for unstable, stable, and neutral conditions using a 1000-m cutoff are 48.4%, 46.3%, and 5.2%, respectively (originally 53.6%, 41.9%, and 4.5%,). Using a 500 m threshold, these percentages change to 44.3%, 44.4%, and 11.2%. We choose to maintain the |L|=1000 threshold because that is consistent with offshore work (Muñoz-Esparza et al., 2012).

Minor Issues

4. **The title needs improvements**

The title suggest that the wake impacts "on" the wind farm development will be studied. This is somewhat inaccurate, as the study is about the wake impacts on offshore wind power production or output, not on the development. Development is choosing the number of turbines or their specs or their layout, which are all fixed in this study; or, development can be how the wind farm installations grow/change with time. Either way, the development here is given (3 scenarios), what changes is the power output.
Also, the title mentions the "Mid-Atlantic" as the focus area, but technically speaking the Mid-Atlantic stops as far north as New York state. From the U.S. perspective, the Vineyard Wind project is not in the Mid-Atlantic and neither are the northeastern lease areas of RI or MA. According to Wikipedia, the following states are included in the Mid-Atlantic: Delaware, Maryland, New Jersey, New York, Pennsylvania, Virginia, West Virginia, and Washington DC. To non-U.S. readers, "Mid-Atlantic" could be the Equatorial zone, as the Atlantic Ocean extends between the two Poles. I don't have a good recommendation for an alternative, but perhaps "U.S." should be added in the title because the study focuses on the U.S. offshore areas after all.
Last, "annual" variability suggests that many years were studied to understand how the production changes from one year to the next. Instead, only one year was simulated here. Thus the variability studied here is monthly/seasonal and diurnal, but not annual.

We keep the siting and characterizations of wind turbines constant in our work and agree that the main focus is on power production. However, our use of Mid-Atlantic is consistent with the Bureau of Ocean and Energy Management terminology. Thus, we will not change this nomenclature. We agree that annual variability implies studying multiple years, and that we should clarify the U.S. focus. We have changed the title to "Seasonal Variability of Wake Impacts on U.S. Mid-Atlantic Offshore Wind Plant Power Production".

5. **Simplify naming**

There is no need to add "_only" to the name of the run with only the Vineyard Wind farm. Just call it "VW."

All instances of "VW_only" have been changed to "VW".

6. **Unclear notation in Eqs. (9)-(10)**

These equations have already been discussed at comment #2, here I am focusing on the notation only. Eliminating "_only" will help (comment #5). P_WV_waked is not defined and uses a notation that differs from that of all other subscripts. All the other subscripts refer to a specific run, whereas "waked" refers to, I believe, a subset of grid points. But the same subset of grid points was used for all other denominators and numerators, thus the confusion. Plus the term P_VW_waked refers to run CA, I believe. I suggest something like (not including my recommendation from comment #2 above):

$$Loss_{ext} = \left(1 - \frac{P_{CA}}{P_{VW}}\right) * 100\%$$

In the text below the equation then you specify that this equation is obtained from the grid cells over Vineyard Wind.

We have changed the notation to reflect the simulation type such that $P_{LA,CA}$ refers to the power production at Vineyard Wind grid cells when exposed to external and internal wakes by either the lease or call areas, to reduce redundancy of writing the same equation twice. P_{VW} refers to power production at Vineyard Wind grid cells in the presence of internal wakes in a VW simulation. P_{NWF} represents power production at Vineyard Wind grid cells from coupling NWF wind speeds to the power curve.

7. **L. 318-320 ("While here ... schemes"): this discussion is irrelevant and unnecessary here.**

This sentence is also redundant and has been removed.

8. **P. 17: some of these results are rather counter-intuitive**

if the TKE_100 runs produce weaker deficits and smaller wake areas, then the wakes should be shorter, whereas the authors report 58 km for TKE_100 and 55 km for TKE_0. The explanation provided is vague and unsupported (l. 346: "larger reduction in momentum aloft"??). The authors do not report exactly how the wake length was obtained. I suspect the method was somewhat empirical and in fact it is giving counter-intuitive results. I suggest that either the authors develop an objective and automated method for calculating the wake length and, if the inconsistency persists, they document and explain it; or that they remove any discussion of the wake length.

We appreciate that, at first, this finding may seem counterintuitive. However, turbulence from the turbines enhances vertical momentum transport from aloft down to within the wake (Gupta and Baidya Roy 2021). The enhanced TKE in a TKE_100 simulation transports more momentum into the waked zone, leaving slower wind speeds above the wind plant (Fitch et al. 2012; Siedersleben et al. 2020). Reduced wind speeds above the turbines then offer a weaker reservoir of momentum available for wake recovery further downwind, leading to slightly longer wake propagation distances. References have been added to support this argument. We explain how the wake length was obtained in Section 2.8 "Wake Identification".

9. **Improve Fig. A1.**

Replace 10-m wind speed with 140-m wind speed in Fig. A1. In all panels (except c), add the results from NWF to appreciate the magnitude of the impacts.

The wind speeds shown in Fig. A1 are indeed hub height wind speed. An error in the y-label lead to a cascading effect in the text which has been fixed. NWF values have been added to the figure.

10. **P. 26-27: the discussion is unclear, the authors report "reductions" in several sentences, but it is unclear what is changing and what the reference is: are they discussing changes from TKE_0 to TKE_100 or from TKE_100 to TKE_0 or from NWF?**

- We now clarify "The wind speed reduction **during this time period** causes a corresponding decrease in turbulent transport of moisture."
- We now clarify "The reduction in heat flux **during this time period** causes 2 m temperatures to decrease and exhibit less variability by TKE amount, with a mean difference of 0.26 K between TKE_100 and TKE_0 (Fig. A1f)."

For the final occurrence of "reduction", it is obvious to the reader that we are referring to the same time period at this point, so to not add distracting information to a topic sentence, we leave this sentence as is: "The reduction in turbulent mixing lowers the PBL, regardless of TKE amount, to shallow heights between 30 to 80 m at 13:00 UTC (Fig. A1e)."

10. **Fig. D1: need a legend for the colors. Also, are these wind speeds or wind speed deficits? The caption indicates wind speed.**

We have added a color bar to Figure D1, included a color bar title delineating that the contours represent the wind speed deficit, and specified the "wake wind speed deficit" in the caption.

Response to Referee #2

Referee comments appear in black and author responses appear in blue.

- line 68: While 12 MW turbines seem to be similar enough to the 13 MW turbines to be installed at Vineyard Wind, I wonder how realistic the assumption is for the other lease and especially the call areas, since those will be build later than Vineyard Wind. Also how sensitive are your results to the chosen turbine type?

We had several discussions with BOEM to determine the best turbine density and the most likely turbine nameplate rating. At that point (Fall 2019), there was little to no knowledge of the actual nameplates to be installed at each lease area, either because it was unknown or proprietary, so a blanket 12 MW was chosen through these discussions.

Repeating these simulations with different turbine types is too computationally demanding and out of scope of this investigation. Other researchers have explored this sensitivity for shorter times periods (Golbazi et al. 2022), finding that the height of the turbine can impact the surface temperature impacts (their Figure 5). A sentence acknowledging this sensitivity has been added to the conclusion: "Further, different sizes of turbines may be installed in some of these regions, and the size of the turbine can influence the impacts of the turbine (Golbazi et al., 2022)."

- line 73 - 74: Why did you choose this period and not a regular calendar year? Do you run continuously or restart the model after a certain period?

We chose this time period due to the availability of lidar measurement data. This clarification has been added: "NWF, VW, and LA simulations run from 01 September 2019 to 01 September 2020 *to capture a full year with available lidar measurement data*". We submitted multiple restarts each month to mitigate runaway error growth as mentioned in new Appendix F.

- line 79 - 81: You don't mention section 3

These are typos where Section "n" incorrectly states Section "n+1," and these have been fixed: "Section 3 discusses variability in stratification, wakes, and power production. Section 4 concludes the work and offers recommendations for future work."

- line 95 ff: Please also provide the WRF option number in addition to the reference

The WRF namelist options for all parameters used in the study are provided in the sample namelist.input, which we provide under the Section "Code and Data Availability."

- Figure 1 caption: last sentence, double mentioning of "red"

The double mentioning has been removed. The last caption sentence now states "E05 (triangle) and E06 (diamond) floating lidars are shown in red."

- line 107 - 114: How realistic is the assumption of regular layout within the areas? To reduce internal wake effects, the turbines might be better placed in an irregular layout

While the goal of minimizing wakes might suggest an irregular layout, minimizing wakes is not the only goal of these wind farms. Cooperative use of these regions requires accommodating other uses. Therefore, our layouts for this work were determined after multiple discussions with BOEM and industry partners. The use of regular layouts in the wind energy areas is realistic, and in fact was requested by other users of this area, notably fishermen and fisherwomen, who request predictable navigable corridors with turbine installations in fixed east-to-west rows and north-to-south columns. For example, https://www.heraldnews.com/story/business/2020/01/07/fishermen-at-odds-with-developers/1945689007/ discusses how a mariner's group supported a regular layout (albeit with even more navigable corridors than proposed here).

- Figure 2a: Where is region 1? Either start numbering at 1 or mention region 1 as below cut-in

The labels for the different regions of the power curve is not something that we developed, but are widely used in the wind energy literature (specifically the controls community) (e.g. Sohoni et al. 2016). We have clarified that "No power is produced in region 1 of the power curve, from 0 m s^{-1} to cut-in wind speed (3 m s^{-1})."

- Figure 3: It would be nice to relate the wind rose to the "regions" in Figure 2. E.g. green could be capped at 11 (below rated power) and one color could be used for region 3. Also "m/s" should be formatted with negative exponents according to the guidelines

Thank you for the suggestion, but we find it is better to retain the granularity in wind speed so as not to limit findings.

- Line 179: Does removing the periods induce a bias? E.g. are they related to the same period / stability category?

Less than 10% of data is removed. The greatest percentage of data is removed during stable stratification, followed by unstable, and neutral conditions at both the E05 and E06 lidars. New table 2 has been added as follows:

Table 1. Percentage of data removed at 140 m due to NaN values.

	Unstable	**Stable**	**Neutral**
E05	1.35%	6.44%	0.33%
E06	3.64%	9.48%	0.62%

- Line 170 - 183: Why do you choose these metrics? How do they compliment each other?
16.440.

The metrics are commonly used in these types of studies. We selected these validation metrics following (Optis et al. 2020), who asserted that these four are key for model-based wind resource assessment. These metrics have been used in subsequent similar investigations (e.g., Pronk et al. 2022). These metrics offer different insight into model performance. For instance, a model may overestimate wind speeds but correctly capture the diurnal cycle, in which case bias would be large but correlation would be strong. Such a setup could present less difficulty for hour-ahead power forecasting, where wind speeds could simply be derated for accurate results. Alternatively, the model could resolve accurate mean wind speeds when compared to lidar measurements but resolve fast wind speeds too frequently. The resulting skewness in the distribution would be captured by the Earth Mover's Distance. Essentially, there are many ways to evaluate if a model is performing well, either temporally, by means, by distribution, etc., and these metrics capture a wide variety of model performance to guide future industry and research decision making.

- Line 189 - 196: Following up on the previous comment, how do you interpret the results that you obtain for the different error metrics? E.g. something along those lines: "the results correlate well in time but have an offset ...". This should also be discussed for the stability based analysis

We provide an interpretation of each metric two paragraphs above where each metric is introduced. The level of detail of this description has been increased: "A CC value of one indicates a perfect correlation between NWF and lidar values. A value of 0 for cRMSE indicates that all values, with model bias removed, lie on the 1:1 regression line. A cRMSE value greater than 0 indicates the distance of residual points from the regression line. Negative biases indicate an underestimation from WRF while positive biases indicate overestimation. A value of 0 for EMD indicates that probability density functions from each data source are equivalent. A positive EMD indicates that the NWF wind speed distribution must shift towards lower values to match the lidar distribution."

- Line 203: You do not describe, which metric you use to classify stability. I assume you are using the same that you use in section 2.7. Consider to move section 2.7 before section 2.6 so that the reader doesn't need to guess.

We agree that it makes sense to build into the validation by providing discussion of the observations, stability classification, and then their combination for the validation. The section order has been switched as suggested.

- Section 2.7: You discuss in Appendix B that the Obukhov length only represents the surface characteristics. Why do you stick to this classification? Also Appendix B should be referenced in section 2.7. Have you estimated the sensitivity of your results to this particular metric? Platis et al. (2021) suggest that depending on the stability metric the results can vary quite a lot (Platis, A., Hundhausen, M., Lampert, A. et al. The Role of Atmospheric Stability and Turbulence in Offshore Wind-Farm Wakes in the German Bight. Boundary-Layer Meteorol 182, 441–469 (2022). https://doi.org/10.1007/s10546-021-00668-4)

We tested the sensitivity of stability metrics between the Richardson number and the Obukhov length and found differences in the percentages of occurrence of unstable, stable, and neutral stratification. We chose the Obukhov length following Archer et al. (2016), who argued that it was a suitable stability metric in the mid-Atlantic offshore region. We have added sensitivity to our choice of a 1000-m cutoff for neutral conditions by adding the percentages of occurrence for each stability class using a 500-m threshold. Also, we have improved the accuracy of the stability metric by calculating the Obkuhov length directly instead of using the WRF-generated values.

"The mean unstable, stable, and neutral percentages of occurrence at Vineyard Wind are 48.4%, 46.3%, and 5.2%, respectively, for the period 01 September 2019 to 01 September 2020, using a 1,000-m threshold for neutral conditions. Using a 500-m threshold for neutral conditions, the percentages are 44.3%, 44.4%, and 11.2%."

- Line 249 - 251: This wake length estimation seems to be too simplified: What about wake turning? What about other wind directions? Arguably the wind rose does show predominant winds from south-west, but other wind directions are also present. In those cases the wake length will be underestimated. To understand your method it would help to draw the line in figure 1.

This wake estimation method compares the wake strength at the same point downwind between unstable and stable conditions, and is consistent with approaches used in the literature (i.e., Rybchuk et al. 2022). Altering the defined downwind line to heterogeneous wake turning or different wind directions would no longer yield a consistent comparison because more factors would be changing than just the stratification.

- Line 270: Reference Appendix E

A reference to the Appendix section (new Section F) has been added: "Power output from VW, LA, and CA simulations are averaged in hourly windows at grid cells containing Vineyard Wind turbines to reduce the effects of numerical noise (Appendix F)."

- Line 304 - 305: This sentence is difficult to understand. Please revise.

This sentence has been revised to "The same pattern occurs elsewhere throughout the OCS because diurnal variability in stratification is weaker than the seasonal cycle".

- Line 311 - 319: These results could be much more neatly presented in a table instead of text form.

Thank you for the suggestion; new Table 3 summarizing the results of this paragraph has been added:

Table 2. Wake wind speed reduction by stratification and TKE amount.

	Unstable TKE_100	Stable TKE_100	Unstable TKE_0	Stable TKE_0
Wind Speed Deficit	-1.5 m s^{-2}	-2.8 m s^{-2}	-1.8 m s^{-2}	-3.1 m s^{-2}
Normalized deficit	16%	25%	19%	27%

- Line 325: "although areal coverage is larger from reduced wind speed replenishment". What do you mean by this?

Because turbulence is weaker in TKE_0, there is less vertical transport of momentum into the waked region from aloft. Accordingly, the spatial extent of wakes grows larger when compared with TKE_100: "although areal coverage of the wake is larger due to weaker turbulence-induced wind speed replenishment from aloft."

- Line 326 - 327: According to the numbers that you present for stable stratification the waked area is actually larger for TKE_100 (16404 km^2) compared to TKE_0 (16060 km^2). This contradicts with your conclusion in this sentence. Please clarify.

Thank you for pointing this out. This sentence has been revised to state that the largest spatial area of wakes occurs in stable conditions in TKE_100.

- Line 341 - 345: Again a table would facilitate a comparison between scenarios

New table 4 has been added underneath the text for easier comparison:

Table 3. The wake wind speed deficit, spatial extent, and downwind propagation distance by added TKE amount.

	Wind Speed Deficit	Spatial Extent	Propagation Distance
TKE_100	-2.2 m s^{-1}	13,040 km^2	43 km
TKE_0	-2.5 m s^{-1}	13,268 km^2	41 km

- Line 349: You reference D1 here, but D1 only shows TKE_100 and thus the differences due to different TKE levels cannot be assessed.

Figure D1 facilitates comparison between stability conditions. This sentence has been clarified accordingly: “The same pattern exists for CA wakes (**Error! Reference source not found.**).”

- Figure 9: Sub-figure titles are (a) for all

Thank you for catching this typo. Sub-figure titles for new Figure 12 have been revised to include (b) and (c).

- Line 361 - 362: Can you provide the power losses averaged over the four month for VW_only and VW_waked for comparison?

The external power losses from the lease areas during the four stable months have been added: “Considering external wakes from the LA at TKE_0 (Eq. 9), the average yearlong power deficit at Vineyard Wind is 14.7% (Fig. 12a) and increases to 15.7% considering only the four stable CA months.” The internal losses over the four stable months have also been added: “During the four CA months only, the deficits increase to 36.9% and 32.9%, respectively.”

- Section 3.3.1: You show also diurnal variations, but these are not discussed. Please add this.

We have added clarification with the following: “While wake-induced losses vary somewhat across the diurnal cycle, there is no discernible pattern. The ocean’s large heat capacity suppresses daytime heating which limits changes in stratification, and by extension, the magnitude of changes in wake losses.”

- Line 383 - 398: It seems a bit counter-intuitively that losses are not additive, i.e. internal losses + external losses != total losses. While the proposed loss estimates (9) and (10) do make sense, they do not share the same reference (P_VW_only vs P_NWF), which makes it more difficult to compare.

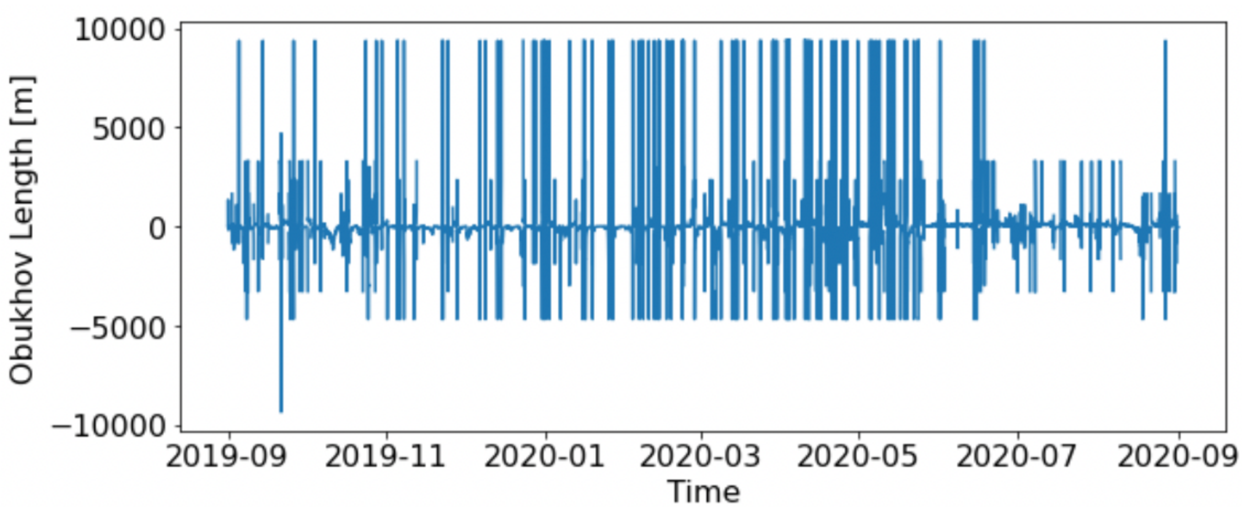
The total wake losses are not additive between internal and external losses, primarily because of nonlinear interactions but also because the denominators are different. We have added an alternative method for calculating external losses, represented as the subtraction between total and internal losses, which share the same denominator in new equation (12).

- Line 402: I understand the energy demand estimates are taken for present day? Are there estimates on how the energy demand will change until CA and LA are build?

New York ISO provides several estimates for future energy demand which vary considerably by the scenario type. A high-load future demand scenario would represent greater implementation of electrification, such as electric vehicles and wintertime heating, and slower adoption of grid independence, such as on-site solar generation. The low-load scenario essentially represents the opposite. (https://www.nyiso.com/documents/20142/2226333/2021-Gold-Book-Final-Public.pdf/b08606d7-db88-c04b-b260-ab35c300ed64). The difference between the high-load and low-load scenarios could reach a spread of about 100,000 GWh by 2053 (https://www.nyiso.com/documents/20142/37320118/2023-Gold-Book-Forecast-Graphs.pdf/ad7db043-ea01-dc3b-b917-ca4cd1d7cd8f). Reporting the amount of demand that the LA and CA layouts could supply in the future would inherit a large amount of uncertainty, which is why we choose to compare supply with current demand.

- Line 408 - 409: Could you add another line in figure 11 representing the stability conditions. This would make it easier to see that the power production is indeed more closely linked to hub-height wind than stability.

Unfortunately, a timeseries plot of stability conditions at the same granularity (as seen below) does not easily facilitate comparison. This is why we have chosen to show stability with longer temporal averages using bar charts and grid plots.



Timeseries of the Obukhov Length over the yearlong period.

- Line 424: Reference figure 2 here again to remind the reader of the definition of region 2 and 3.

A reference to the power curve in Figure 2a has been added: " These differences are small at slow wind speeds, because little momentum is available for wake recovery, and at faster wind speeds within region 3 of the power curve (11−30 m s^{-1}) where wind speed changes do not affect power production (**Error! Reference source not found.**a)."

- Figure 13 caption: "black dots indicate turbine locations": suggesting to add "in TKE_0 and TKE_100", since in NWF they are not included

The caption (for what is now Figure 16) has been rewritten to "black dots indicate turbine locations in VW TKE_0 and TKE_100".

- Line 508: It would be interesting to discuss, how the difference due to added TKE amount compares to the difference due to different PBL schemes. You mention Rybchuk et al. (2022) at some places through the paper, but don't compare the effects due to PBL schemes and added TKE amount directly.

While we would also find this an interesting discussion, direct comparisons are not possible as Rybchuk et al. focus on idealized scenarios and the present study is for real scenarios. We are currently working on winning funding to carry out a more direct comparison of the present work with simulations with the 3DPBL scheme. Detailed discussion is not within the scope of this work, and so we refer the readers to Rybchuk et al. (2022).

- Line 537: What do you mean by "the differences ... are precise"?

This sentence has been rewritten to "The sequencing of power production driven by TKE amount remains consistent, namely that the differences always progress from TKE_0 to TKE_25 to TKE_50 to TKE_75 to TKE_100." "Consistent" is used as a lead into the next sentence where we discuss that power production values are typically bookended by TKE_0 and TKE_10.

- Appendix A: The mixture of discussion on variability due to added TKE amount and the special case during calm winds between 12:00 and 15:20 on 12 July is confusing. These two aspects should be kept separate.

We attempted to rewrite the section on the special case-study period separately as the reviewer suggested, but upon reading, we determined it was even more confusing to bounce back and forth between meteorological variables (wind speed, heat flux, etc), and decided that in this appendix, we will keep each idea in its own respective paragraph.

- Line 555: the first sentence is a bit difficult to understand. The difference between TKE_0 and TKE_25 seems to be more than 15 to 20 m

15-20 m here refers to the actual (very shallow) boundary layer height for a specific time period and not the difference between runs. This sentence has been rewritten to "The reduction in turbulent mixing lowers the PBL, regardless of TKE amount, to very shallow heights between 15 to 20 m from 12-15:20 UTC (**Error! Reference source not found.**e)."

- Line 565: The way you reference figures is sometimes confusing to me. For instance, I would reference Figure B1 here as "stratification at the E05 and E06 (Fig. B1) lidars exhibits similar seasonal variability to Vineyard Wind (Fig. 6)". Since vineyard wind is shown in Fig. 6 and not in Fig. B1. Please also check other parts of the manuscript. Note also that you wrote "E05" twice.

Thank you for pointing this out. A figure reference should go directly after the point being made. This recommendation has been implemented.

- Figure D1: Colorbar is missing; is the upper row just a zoom of the lower row?
Yes, it is just a zoomed in version of the figure. We have added the explanation that "The upper row is zoomed in to increase granularity" in the figure caption. A colorbar has also been added.

- Figure E1: "at which the map occurs" -> suggestion "of the map"

The caption text has been revised according to the suggestion in new Figure F1: "The gray vertical line shows the time stamp of the map."

- Line 643 - 646: Difficult to understand. What do you mean by "poses a threat to power estimations". I don't understand the contrast "although ..., we show noise occurring in the SE ..." and why this "underscores the point that ... should only show differences within the wake". Please clarify.

- The first sentence has been clarified to: “Noise occurring in grid cells containing turbines could undermine power estimation accuracy and we observed noise occurring in the southeastern portion of the domain.”

- The second sentence was changed to be more concise: “Subtraction of wind speeds between simulations with variable TKE amounts should only show differences within the wake, and such differences are a result of noise.”

- Line 660: Is there a link missing for "OpenEI_link"?
We are still working on getting the data ported for public access. A url will be inserted here once the data is uploaded.

- Line 715: Missing DOI
A DOI has been added

- Line 717: Missing URL
A url has been added

- Line 839: Missing URL
A url has been added

- Line 844: Missing DOI
A url has been added

Response to Mark Stoelinga

Referee comments appear in black and author responses appear in blue.

We thank the reviewer for his thoughtful comments.

181-182: I think centered RMSE (cRMSE) is essentially the same as what I’ve heard and referred to as bias-corrected RMSE (or BCRMSE), in which you first calculate the mean model bias error, subtract it from all the model values, then calculate RMSE. And, I believe both are essentially equivalent to the standard deviation of the errors as well. All that is neither here nor there. However, I do think the sentence in lines 181-182 should be clarified, to say that “a value of 0 for cRMSE indicates that all values, ***after removal of the respective model or measured means***, lie on the 1:1 regression line”.

Thank you for the suggestion. We have added your proposed clarification to this sentence as “A value of 0 for cRMSE indicates that all values, with model bias removed, lie on the 1:1 regression line”.

189 (paragraph): Might be good to show model versus measured mean shear exponent, a metric that the wind industry uses extensively and is highly familiar with its typical range of values.

Thank you for the suggestion. We have added a new Figure 8 and a discussion of model versus measured wind shear exponent, finding that lidar-derived exponents are in good agreement with past evaluations in the mid-Atlantic and that WRF-derived exponents are underestimated.

325-326: There is an interesting result in Fig. 8 that you do not comment on, which is similar to behavior other have seen and commented on (including, I believe, one or more of you in previous work, and myself). What I'm referring to is the opposite effect of TKE amount in the near-project versus distant wake environment. Within and near the project, behavior is intuitive: higher TKE dissipates wakes and leads to smaller waked wind deficits. However, farther away, as evidenced by the distance northeastward of the first (0.5 m/s) contour, as well as the area of this contour reported in the text, it is actually slightly farther (and covers more area) with TKE than without it. In other words, at distance, higher TKE actually helps wakes, whereas near or within the project it hurts wakes. I saw the same behavior, and I'm certain you and others have commented on it previously. Do you have any new insights into this behavior?

This comment was clarified and retracted by the reviewer in a later comment posted in the online discussion.

Appendix E. The authors and I have had discussions in the past about the nature of the noise seen in difference fields (turbines minus no turbines wind speeds). I'm not opposed to the idea that they are purely numerical; I agree that is the most likely explanation. However, I still consider it possible that even the distant differences are perhaps partly physical rather than numerical. They tend to occur in an unstable boundary layer or in convective scenarios. These scenarios are characterized by small-scale, high-amplitude, chaotic structures (convective cells) whose initiation locations are random and probably sensitive to even the smallest perturbations, which may include very subtle and fast-moving gravity waves or other disturbance triggered by the presence of the turbines. For the purpose of energy production, though, they are probably inconsequential because they tend to cancel each other out when averaged either spatially or temporally.

Apart from noise adjacent to the farms, we have observed noise also appearing far upwind of the turbines where the introduction of wind plants should make no discernible difference to the atmospheric state (tens of kilometers upwind of the induction zone, with little or no noise in the induction zone). Even if gravity waves were involved here, gravity wave deflection should maximize close to the wind plants before dissipating, making it more likely that these features are numerical, but we agree that numerical noise is worth looking into in future studies.

References

Archer, C. L., B. A. Colle, D. L. Veron, F. Veron, and M. J. Sienkiewicz, 2016: On the predominance of unstable atmospheric conditions in the marine boundary layer offshore of the U.S. northeastern coast. *Journal of Geophysical Research: Atmospheres*, **121**, 8869–8885, https://doi.org/10.1002/2016JD024896.

——, S. Wu, Y. Ma, and P. A. Jiménez, 2020: Two Corrections for Turbulent Kinetic Energy Generated by Wind Farms in the WRF Model. *Monthly Weather Review*, **148**, 4823–4835, https://doi.org/10.1175/MWR-D-20-0097.1.

Fitch, A. C., J. B. Olson, J. K. Lundquist, J. Dudhia, A. K. Gupta, J. Michalakes, and I. Barstad, 2012: Local and Mesoscale Impacts of Wind Farms as Parameterized in a Mesoscale NWP Model. *Monthly Weather Review*, **140**, 3017–3038, https://doi.org/10.1175/MWR-D-11-00352.1.

Golbazi, M., C. L. Archer, and S. Alessandrini, 2022: Surface impacts of large offshore wind farms. *Environ. Res. Lett.*, **17**, 064021, https://doi.org/10.1088/1748-9326/ac6e49.

Gupta, T., and S. Baidya Roy, 2021: Recovery Processes in a Large Offshore Wind Farm. *Wind Energy Science Discussions*, 1–23, https://doi.org/10.5194/wes-2021-7.

Muñoz-Esparza, D., B. Cañadillas, T. Neumann, and J. van Beeck, 2012: Turbulent fluxes, stability and shear in the offshore environment: Mesoscale modelling and field observations at FINO1. *Journal of Renewable and Sustainable Energy*, **4**, 063136, https://doi.org/10.1063/1.4769201.

Optis, M., N. Bodini, M. Debnath, and P. Doubrawa, 2020: *Best Practices for the Validation of U.S. Offshore Wind Resource Models*. National Renewable Energy Lab. (NREL), Golden, CO (United States),.

Pronk, V., N. Bodini, M. Optis, J. K. Lundquist, P. Moriarty, C. Draxl, A. Purkayastha, and E. Young, 2022: Can reanalysis products outperform mesoscale numerical weather prediction models in modeling the wind resource in simple terrain? *Wind Energy Science*, **7**, 487–504, https://doi.org/10.5194/wes-7-487-2022.

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

Siedersleben, S. K., and Coauthors, 2020: Turbulent kinetic energy over large offshore wind farms observed and simulated by the mesoscale model WRF (3.8.1). *Geoscientific Model Development*, **13**, 249–268, https://doi.org/10.5194/gmd-13-249-2020.

Sohoni, V., S. C. Gupta, and R. K. Nema, 2016: A Critical Review on Wind Turbine Power Curve Modelling Techniques and Their Applications in Wind Based Energy Systems. *Journal of Energy*, **2016**, e8519785, https://doi.org/10.1155/2016/8519785.