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 km²) compared to TKE_0 (16060 km²). 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.

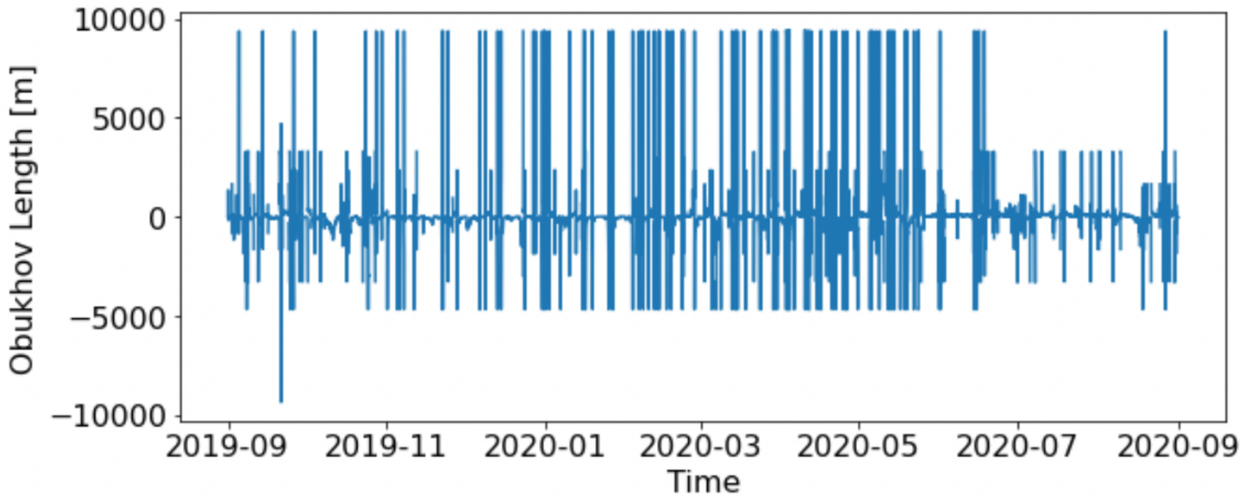
The total wake losses are not additive between internal and external losses, primarily because of nonlinear interactions but also because the denominators are different. We have added an alternative method for calculating external losses, represented as the subtraction between total and internal losses, 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

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.

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.

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, 2021: The Sensitivity of the Fitch Wind Farm Parameterization to a Three-Dimensional Planetary Boundary Layer Scheme. *Wind Energy Science Discussions*, 1–39, https://doi.org/10.5194/wes-2021-127.

——, ——, ——, ——, ——, and ——, 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.

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.