Response to Referee Comments for "Measurement-Driven Large-Eddy Simulations of a Diurnal Cycle during a Wake Steering Field Campaign"

Thank you to all of the referees for your time and thoughtful feedback, as well as your kind comments. The updated manuscript is a significant improvement upon my original submission in terms of clarity and completeness. Major changes to address similar comments from all three referees include:

1. Abstract: The statement, "… an LES with actuator-disk turbines was compared to a steady-state engineering wake model, demonstrating agreement with measurements under partially and nearly waked conditions" is too general and errors should be reported.

    o Mean absolute error (MAE) has been reported for the final simulated atmospheric conditions that were used as input to the turbine simulations: "During the selected turbine-analysis periods (with and without wake steering), the simulated wind speed ranged between 4.7 and 7.1 m s^{-1}, the wind direction was approximately north–northwesterly, and the TI ranged between 0.02 and 0.08. The mean absolute errors (MAEs) for these three quantities were 0.19 m s^{-1}, 1.5°, and 0.031, respectively." (lines 375–378).

    o The results sections (Sect. 5.1 and 5.2) have been cleaned up. Analysis period labels have been added to Figs. 9 and 10 to facilitate the discussion in the text, which has been updated accordingly. All averaged quantities are now plotted at the center of each 10-min analysis period (previously they were at the left edge of each 10-min interval) to facilitate interpretation. To reduce ambiguity when reporting errors in TI, I have switched from % to being unitless.

    o All analyses have been repeated and numerical quantities verified to match what is plotted in the figures. During all the analysis periods, I now report statistics about the wind speed, wind direction, and TI error.

    o For the two most representative analysis periods (periods "3" and "4"), i.e., when the up- and down-stream turbines are closest to being aligned and the yaw offset is ~0° and a maximum of 15°, I have reported errors in normalized power.

    o Finally, the abstract has been updated with:
    "Subsequent analysis identified two representative periods during which the up- and down-stream turbines were most nearly aligned with the mean wind direction, and had observed yaw offsets of 0° of 15°. Both periods corresponded to partial waking on the downstream turbine, which had errors in LES-predicted power of 4% and 6%, with and without wake steering."

2. Appendix D: It became readily apparent that inclusion of the mesoscale results was essential to putting this work in the right context. The WRF model setup, results from

simulation ensembles during the wake-steering case study, and tabulated errors are included.

3. The original manuscript lacked clarity about why the selected MMC methods were evaluated. I have added a new section (4.1) that reviews applicable methods prior to the results section.

4. To provide additional justification for why I used the partial profile DPA approach, I have updated Figure 8 and added discussion about the simulated and measured wind shear in Section 5.1. I think this has made the MMC results for the stable boundary layer more interesting because it appears that partial profile approach successfully captures a pair of shear instability events in the nighttime, just prior to the case study periods.

I have also updated the paper acknowledgements and the reference to the companion paper:

Shaler, K., Quon, E., and Jonkman, J.: Wind Farm Structural Response and Wake Evolution for an Evolving Stable Boundary Layer: Com- putational and Experimental Comparisons, Wind Energ. Sci. Discuss., in review, https://doi.org/10.5194/wes-2023-138, 2023.

Please see below for responses to comments from each referee.

Referee 1, "Inflow is all you need"
https://doi.org/10.5194/wes-2023-101-RC1

> *Nice effort to simulate a realistic ABL evolution during a wake steering experiment. Capturing the interplay of the inflow with the wake dynamics is a notable challenge and the author made a good job addressing some of the most important couplings with high-fidelity and engineering models.*
>
> *Getting the inflow conditions right is the main challenge and it is there where I have focused my attention while reviewing.* **Different profile assimilation techniques are discussed but it is not justified why we need to add such complexity to a case study that seems to be dominated by surface-layer conditions and single-point assimilation may be a more effective solution to track mesoscale variability.**

Throughout the revised manuscript, I have attempted to clarify how and why I used the different assimilation techniques. In this revision, I have also included an analysis of wind shear evolution in the nocturnal SBL, which further motivates the partial-profile direct profile assimilation approach. More details are provided in my response to line 403 below.

> *Another topic of interest is* **to which extend we can extract some quantification of the impact of mesoscale inflow bias at the reference meteorological site and the bias due to horizontal heterogeneity not simulated explicitly by the models**. *By looking at the normalized power in T2, it is observed that these biases can lead to up to 20% differences in gross power. A fairer assessment of the wake models would be obtained if those biases could be quantified to gage their relative role in the assessment of wake steering.*

I have now included my preliminary WRF results in Appendix D, which shed some light on the mesoscale inflow bias. Fully assessing the bias due to horizontal heterogeneity would require a terrain-resolving LES on a larger domain and is beyond the scope of the present work. I agree that a fairer assessment of the wake models would be obtained if we could better quantify uncertainty. As prompted by Referee 3, I have also added some additional comments in the discussion and conclusions that speak to the uncertainty associated with dynamic inflow effects.

> *In summary, I think it is a good case study to highlight the challenges of simulating realistic operational conditions departing from the traditional canonical microscale setup. Still, **the large uncertainties in the inflow conditions make it extremely hard to isolate the assessment of wake steering effects. Maybe we need to change the orientation of profile assimilation techniques from the vertical to the horizontal direction**. We need the inflow bias to be in the order of a few percent of normalized power to study wake steering effects that can stand out of this bias.*

I agree that for wake steering effects to stand out our model errors need to be within a few percent, and getting the power right requires getting the inflow right. The engineering model (FLORIS) appears to be sufficiently accurate for bracketing the downstream turbine performance in the extremes of completely waked or unwaked conditions, at least for two turbines. What this study shows, I believe, is that there are some interesting and challenging intermediate cases: partially waked (in which LES seemed to do reasonably well and better than FLORIS) and nearly/slightly waked (in which LES appears to capture the appropriate flowfield physics, but the accuracy may be more sensitive to inflow characteristics).

It's an interesting idea to consider horizontal assimilation techniques but would require additional computational effort to solve on a finite aperiodic domain. A strategy for setting the source terms based not on planar averages but sampled point locations would be needed, as well a way to distributed the forcing terms in three dimensions. Happy to discuss this idea further.

Additional remarks

12: "demonstrating agreement with measurements": I'm not sure I would agree with such a general statement. Can you be more specific or also mention where there is disagreement?

- The other referees also raised similar concerns. Please see the general response at the beginning of this document.

22: Nonstationarity is intrinsic to atmospheric flows so it really happens all the time although it is more notable in the presence of weather events and diurnal morning/evening transitions as the authors indicate. Maybe it would be worth speaking about "large-scale nonstationarity" in the context of this paper, to differentiate from the more traditional microscale nonstationarity that leads to quasi-steady canonical conditions, a.k.a. "microscale LES".

- Thank you for this suggestion. I have added a comment in the introduction, "Nonstationarity occurs across a range of scales, from quasi-steady canonical conditions

in the microscale to synoptically-driven atmospheric motions at large scales." (Lines 26-27)

92: Fleming et al (missing year)

- Thank you for pointing this out, this has been corrected.

115: Can you describe how you calibrated the turbulence intensity of the lidar from the cup anemometer measurements? How far off where the lidar measurements at the calibration levels?

- The third referee also wanted more detail here. I have added a bit more detail (lines 125–128), based on previous internal communication with the colleague that had provided the lidar data. That colleague has since left NREL and, as far as I know, has no plans to publish his work on the lidar turbulence correction. I have also calculated and included summary statistics based on archived data.

121: Can you provide an estimate of the uncertainty of the yaw position achieved after applying the corrections to the SCADA signal?

- I unfortunately could not find any information on the uncertainty in the yaw measurement. To be clear, we installed our own yaw encoder and I did not use the SCADA yaw signal for any analysis.

220: Why using sounding data from 340 km instead of local mesoscale predictions? I wonder if the authors checked the accuracy of the mesoscale data at the sounding site. This could give some guidance as to whether the mesoscale data could be used reliably instead of relying on distant measurements. Would be nice to include the mesoscale profiles in Figure 4 to get a sense of the differences in this indirect validation.

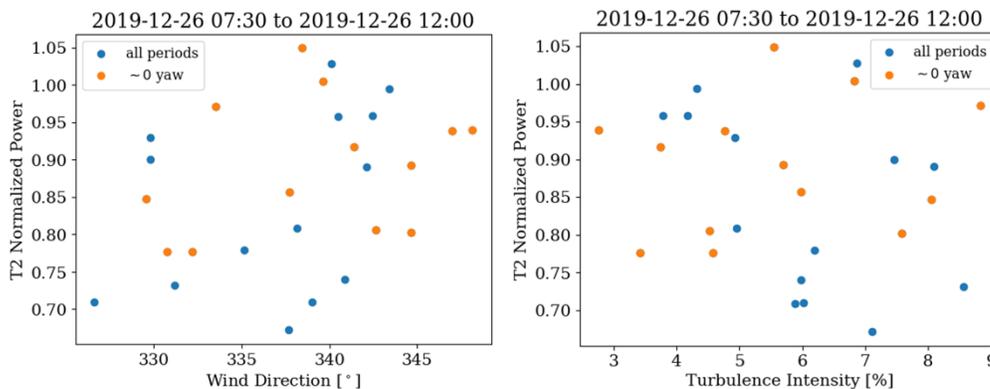
- Please see the new Appendix D for an assessment of mesoscale model accuracy at the test site.
- The distant soundings could have been used to validate the mesoscale predictions during the study period but this was not done, considering the poor comparisons between mesoscale model predictions and local measurements. A hybrid approach could have also been used, blending WRF predictions aloft with local measurements in the surface layer. However, I decided early on to abandon mesoscale modeling altogether and attempt to simulate local mesoscale variability using only available measurements.

336: There are notable differences of up to 20% in the power of T2 compared to the average from freestream turbines T1 and T5 in conditions without wake steering control. Is this not an indication of predominant heterogenous inflow? It is very challenging to assess potential net gains of a control strategy of a few percent when one of the hypothesis of the model can lead, by itself, to differences of up to 20%.

Similarly, one could try to estimate the impact of the differences observed in the reference inflow conditions (Figure 9), especially regarding wind direction differences which result in an indirect

yaw offset. Is it possible to infer from measurements or simulations the impact of these inflow biases in the freestream power?

These estimations, in the form of power ratios, would provide a more comprehensive framework to judge the relative importance of each phenomena individually so we can make more informed conclusions about their relative role when all factors are combined.

- I agree that there is a lot of uncertainty due to inflow heterogeneity and it makes the results challenging to interpret. I have viewed this study through the lens of "if we simplify the problem, can we still get close to the right answer?"
- Per your suggestion, I checked to see if there is a relationship between wind conditions and the normalized power of T2. There unfortunately does not appear to be any clear relationship between power and wind direction or turbulence:

At the moment I do not think I can come up with an estimate of any ratio that can characterize the heterogeneity or correct for it.
- I have also added additional comments in the discussion describing this: "The effects of heterogeneity can also be clearly seen in the variability in power produced by the upstream turbine T2 (Fig. 10). No relationship was found between the measured power from T2 and measured wind quantities in Fig. 9." (Lines 464–465)
- A follow-on simulation study, involving resolved terrain and turbine T2, for less than an hour (e.g., 0730 to 0820) could shed some light on the effect of heterogeneity. Inflow conditions could be provided by the present precursor simulation and comparing the two simulations could reveal some local flow bias. However, I feel that would constitute more than just another section of this paper and deserves to be its own body of work. It would bring new challenges, e.g., how much resolution is needed to resolve important terrain-induced flow features or the validity of Monin–Obukhov Similarity Theory over nonflat terrain.

403: Concerning the data assimilation techniques it is not clear how the authors ended up choosing the DPA method over the much simpler single-level method. From Figure 8 it seems that the single-level method produces a more accurate representation of the friction velocity during the wake steering experiment while it shows similar performance than the other models in the other quantities. Maybe the case is not large enough for tall profile features to play a role and

the flow is dominated by the bottom part of boundary layer, which can be calibrated with a single point to capture the mesoscale tendencies.

- I agree that the original manuscript did not make clear why I ended up choosing the partial-profile DPA approach over the simpler single-level method. I have updated Figure 8 to include the resulting wind shear, which I feel makes a compelling case for partial DPA in the nocturnal stable boundary layer.
- You raise a valid point that the friction velocity results seem to suggest that the single-level method is best and the other methods miss the mark. I realized that I neglected to comment on this in the original manuscript and have added a sentence at lines 364–366: "In the early morning after 01:00 LT, the friction velocity at 10 m is only captured by the single-level assimilation. Considering that this occurs during stable conditions, this discrepancy may be a consequence of inadequate grid resolution near the surface (10 m, see Appendix B for additional LES details)."

Referee 2
https://doi.org/10.5194/wes-2023-101-RC2

> *In this study, the author did a great job discussing and addressing some of the most challenging aspects of MMC couplings. Ability to accurately simulate conditions wind turbines are exposed to, especially in case of a complex terrain, persists as a challenge for the wind energy community. Better documented profiles, appropriate assimilation technique and accurate inflow conditions are essential component in addressing those challenges and reducing uncertainly that can lead to a significant bias in projected power production. I think study is a great example of challenges one would face when simulating more realistic ABL conditions, and those that wind energy community needs to overcome.*

Specific Comments

Line 12: "agreement with measurements" based on presented results, I feel that author should rephrase this a bit.

- The other referees also raised a similar issue. Please see the general response at the beginning of this document.

Line 92: "Fleming et al" here is missing a year.

- Thank you for pointing this out, this has been corrected.

Line 220: Author used sounding that is 340 km upstream from the study location. How far is second available sounding for the same time period and would it be useful to do some sort of spatial interpolation or comparison between those location for a possible better informed profiles of wind speed and virtual potential temperature? I understand that WRF was not able to properly

simulate specific local conditions, but was WRF able to properly simulate wind and temperature profiles at some location closer than 340km (location of the closes sounding), and would it be better to use some of those closer to domain location profiles?

- The nearest soundings were all launched over 100 km from the test site. While it is certainly possible to interpolate in space between sounding sites, I believe this would have even more uncertainty than the WRF mesoscale solution. Please see the new Appendix D for an assessment of the mesoscale model accuracy.

Line 326: Maybe I missed it, but what was a rational of using partial-wind-profile DPA compared to full-wind-profile DPA?

- Referee #1 also had a similar question. As discussed in lines 472–476, the partial-profile DPA is able to capture a realistic nocturnal low-level jet as well as shear instabilities that were observed. I have also updated Fig. 8 to include a time history of the mean wind shear to showcase the latter point.

I highly agree with the statement that "the extent to which field conditions are reproducible with MMC depends on the nature of the background physical phenomena and their observability" and think that this work is a very good example of how tools and techniques presented in this study can improve prediction of local features and conditions only for a limited extend.

- Thank you very much for this specific feedback.

Referee 3
https://doi.org/10.5194/wes-2023-101-RC3

> *The submitted paper investigates an approach to simulate a published wake steering field campaign in large eddy simulations (LES) that derive forcing from the limited field measurements available. Overall, this paper is very valuable to the community, and an impressive and detailed effort to simulate more realistic atmospheric conditions in microscale LES.*
>
> *The paper is overall well written, and includes a lot of detail on certain aspects, but more limited details elsewhere, given the nature of applying existing methods and approaches to a given case study. My overall perspective is **that this paper would be both easier to follow, and in my opinion more impactful to the community beyond users of NREL codes, with a restructure. Specifically, I'd like the methods section to detail meso-microscale coupling (MMC) methods that could be used in general, and for the results to intercompare the chosen methods.** As is, Section 4 introduces one MMC method that has been seemingly preselected as the best, but then four different MMC methods are evaluated in results, with a 5th method being WRF-LES MMC, which was simulated but is not shown apparently because of poor performance (would be better to show). I think if the author approaches the paper as an intercomparison, with a discussion of benefits and limitations of each method, it would be easier to follow, and also useful to the community.*

- I apologize for the confusion but I never ran "WRF-LES MMC", which I would interpret as being WRF nested down through the terra incognita to LES scale. I'm not sure which part of the text may be revised to clarify this. However, the mesoscale WRF solution is now included in Appendix D.
- I agree that readers unfamiliar with the previous work in profile assimilation may be thirsting for more information. An introductory paragraph has been added to the beginning of Section 4. Section 4.1 has been restructured, with references added to Haupt et al 2023 (line 220), which provides an overview of MMC methods, and Allaerts et al 2020 (line 240–241), which compares the direct and indirect methods.
- I have also moved some of the text from Section 5.1 to Section 4.1 and stated up front that I am evaluating all applicable MMC coupling methods. I hope that it is clear I am not preselecting a particular method—Section 5.1 shows an intercomparison. While the partial DPA approach was selected for the nocturnal stable boundary layer and wake steering study, during the daytime the preferred method was in fact the simpler single-level approach (lines 334–337).
- Lastly, to help clarify the methodology for a broader audience, I have added Table 2 to summarize the various MMC approaches.

*Further, there are a lot of choices in the simulation setup. This is not a criticism, in fact, this is part of what makes this paper useful and impressive. But somehow, **it would be nice to very clearly and concisely summarize the choices made** (not a complete list: guessed roughness length, potential temperature B.C., initialization with sounding data from far away ad hoc combined with in situ temperature measurements, spin-up period, power-law fitting, actuator disk model, Boussinesq approx., horizontally homogeneous flow, MMC method, etc.). Many of these choices were selected based on intuition, some were justified based on empirical results, and some were tested in sensitivity tests. It would be really helpful to just clearly list this, **discuss why each decision was made in a concise summary, and to highlight the need for future work. If someone else wanted to start where this paper ended, why should they make similar or different decisions on the setup***?

- A new table has been added to the appendix (Table B1) summarizing all MMC modeling choices in one place.
- I have added some additional details: the use of the Boussinesq approximation (line 533) and elaborating on the choice of roughness height (lines 278–282).
- I have also added an additional paragraph to the Discussion Section 5.3 (lines 477–487) with suggestions for future directions.

*Finally, I feel that **the 'final' wake steering results in Figure 10 are somewhat unconvincing, due to the multitude of choices and uncertainties discussed above, and due to the short term simulation window.** It seems hard to make any conclusions from these data, as I listed below in the point comments.*

- Referee 1 also had a similar concern. For this study, I have attempted to clearly document all the complexities of the case study and acknowledge all of the limitations and

implications of my assumptions. Ultimately, I have viewed this study through the lens of "if we simplify the problem, can we still get close to the right answer?"

- I did some follow-on analyses to try to characterize or correct for the heterogeneity or dynamic inflow effects. However, they were wholly inconclusive. To better understand these results, a longer simulation window could be considered especially if modeling loads is of less interest (that was the most restrictive filter on data availability); a terrain-resolved LES could also be used to understand the effects of terrain on the mean flow and turbulence intensity. However, I believe that study would constitute its own body of work.

Point Comments

1. **Line 73**: Can you clarify what is intended by this statement *"Not only is NWP not able to resolve local conditions, the available remote-sensing observations are limited."*

   It seems to me from the detailed literature review in the preceding paragraphs that the application of interest in this study would be well-suited for MMC using NWP for the large-scales, if in situ measurements are limited.

   - The paragraph at lines 78–88, "The objective of the work discussed herein…," has been rewritten for clarity. In particular, lines 82–85: "The identified case study precludes straightforward usage of previous MMC strategies for two reasons: 1) NWP is unable to resolve local conditions so that microscale atmospheric forcings are necessarily derived from measurements and 2) local measurements of the wind and temperature profile are not continuously available over time and do not span the entire height of the computational domain."

2. **Line 100**: Please clarify the filtering/time-averaging associated with the wind speed and direction that feed into the lookup table.

   - Please refer to Simley et al 2020 (added to line 109) and Fleming et al 2020 for details about the lookup table design.

3. **Line 115**: *"Lidar-measured turbulence intensity above the 60-m-tall met mast was corrected based on lidar and cup-anemometer measurements."*

   Could you describe the technical details of this more specifically?

   - Referee #1 also had a similar question. I have added a bit more detail (lines 125–128), based on previous internal communication with the colleague that had provided the lidar data. That colleague has since left NREL and, as far as I know, has no plans to publish his work on the lidar turbulence correction.

4. **Section 2.2**: I wonder if this section and associated discussion elsewhere in the paper, can be reorganized in three parts: 1) what input data are required by the MMC approach; 2) what data are available during the field campaign; 3) what methods were used to fill

gaps/missing/imperfect information. To be clear, I thought point 1 (what data are required for your specific MMC approach) was most unclear in the study.

- Please refer to Table 1 for a summary of available data.
- A new Table 2 has been added, which compares different applicable MMC strategies given the modeling assumptions in this study. Typical use cases and expected output behavior is included.

5. **Section 2.3**: What is the maximum time period that could be simulated given computational constraints? The reason I am asking, is because given all of the filter constraints, only one hour of data is quite limited.

- I am not sure where the referee has identified the one hour of data but I agree that the amount of data that met the filtering criteria was quite limited. Ignoring the companion study that focuses on loads would have made a full year available for consideration (and perhaps resulted in the selection of a more benign case).
- The maximum time period that can be simulated would depend on the availability of computational resources and the goal of your computational study.
  - It should be possible to simulate the ABL for the full two months of the field campaign that included loads instrumentation. SOWFA is built on OpenFOAM, which is not a highly performant code. For the full simulation day, SOWFA required 52 h of wallclock time while running with 504 CPU cores.
  - In comparison, simulating the 4.5 h of the wake steering study including actuator disks (as well as spatial and temporal refinement) required 48 wallclock hours on 864 CPU cores.
  - Newer solvers, especially if able to make use of GPUs, will see much faster turnaround times.

6. **Line 156**: This sentence self-references the section that the sentence appears in

- Thank you for catching this. The original sentence has been removed and the paragraph containing the original sentence has been rewritten for clarity (lines 165–176).

7. **Section 3.1**: I suggest including an Appendix that contains the mesoscale modeled results and associated comparisons with measurements to support this discussion.

- A new section, Appendix D, has been added. It details the mesoscale model setup and compares results with measurements (lidar hub-height wind speed and met-mast 2-m temperature) for different initial and boundary conditions. Mesoscale model errors during the case study are quantified, generally showing a wind-speed overprediction of 4–5 m/s and a wind-direction offset of 13–17° (Table D2).
- Appendix D has been referenced in Section 3.1, at line 165. I no longer have the archived results for different PBL schemes, so I have excluded that from the list of covariates in the sensitivity study (lines 166–169).

8. **Section 3.2**: It seems to me that the filtering approach discussed in Section 2.3 has resulted in a set of complex meteorological conditions. Not a positive or negative per se, but maybe it would have been useful to start with a simpler case before moving to a more complex one.

- I completely agree. However, a very simple case (diurnal cycle over flat terrain in Texas) was already previously considered (the Allaerts et al. papers) and an objective of this study was to enable a high-fidelity study of loads during the wake steering campaign. If this case can be considered very challenging, and a more "moderately challenging" case would have been ideal.

9. **Line 183**: *"The third and final assumption for this case study [...]"*

It may be easier to read for most if you state this assumption in terms of the standard Boussinesq approximation.

- I do not mean to directly reference the Boussinesq approximation here. For clarity, I have rewritten the sentence, which now reads: "The final assumption for this case study is that for wind-energy quantities of interest, the transport of moisture (seen in Fig. 3) does not need to be explicitly modeled. Instead, the effect of moisture on the total air density is implicitly captured through virtual temperature quantities." (lines 211–213)

10. **Line 219**: Is this a typo? The sounding data are 340 km away? Why were these data used instead of the pressure, temperature, and humidity sensors mentioned in Table 1?

- No, this is not a typo. The upstream sounding data were used to provide an estimate of the upper-air conditions, out of range of the instrumentation in Table 1. My expectation was that the wind and temperature at higher altitudes should change over larger spatial scales and longer time scales and would therefore still be representative of upper-air conditions at the test site. The local sensors were then used to correct the remote profiles near the surface, as shown in Figure 4.
- The discussion at lines 252–253 has been updated for clarity, "Because information about the upper atmosphere is not available from local measurements, …"

11. **Line 227**: *"To adapt the nearest sounding to local conditions, the lowest 200 m of the wind and virtual potential temperature profiles were replaced with local site measurements"*

Did this sharp discontinuity in the field cause any challenges or instabilities? Are the results sensitivity to this choice, rather than the alternative approaches of using only the sounding or using only the local measurements? Is the capping inversion in Figure 4(c) only the result of the jump between local and sounding measurements?

- These are all good questions. The discontinuities in the initial field did not create any simulation challenges.
- Another preliminary study I performed, using only local measurements and informed guesses about conditions aloft, showed sensitivity to both the height and strength of the capping inversion. I therefore could not simply choose reasonable values from literature and past experiences. However, it should be noted that the differences between the simulated turbulence realizations were more pronounced in the daytime than at night in the stable ABL. Given the number of uncertain quantities in this study, I elected not to have the LES initial conditions be a free parameter in my study.

- The capping inversion in Figure 4c comes from the sounding data. Please note the small symbols on the blue curve indicating the actual measurements and the temperature gradient that was captured.

12. **Line 243**: I didn't quite understand this last line in the context of the paragraph. If I'm reading correctly, it says to use surface temperature instead of heat flux (Basu 2008), but then heat flux is used as the boundary condition for a reason I didn't quite follow

- Thank you for reading this section very carefully. I have rewritten this paragraph (lines 268–277) for clarity and removed irrelevant statements about temperature profile assimilation.
- I also noted in the Discussion Section 5.3 (lines 484–485) that specified temperature surface conditions could have been used.

13. **Line 250**: Couldn't this discrepancy also be associated with a mis-specification of the geostrophic pressure gradient?

- The geostrophic wind or background large-scale pressure gradient is actually specified by the momentum source term that is dictated by the MMC approach. I expect that changing this forcing term will change the surface shear stress and, consequently, the atmospheric stability. This in turn would affect momentum and temperature fluxes in the surface layer, so I believe your reasoning is valid. However, there is no mechanism to dial up or down this pressure gradient without modifying the quantity being assimilated. I believe the simplest way to correct for the mismatch I noticed was to adjust the surface flux as I have done.
- Modifying the geostrophic forcing is essentially what the partial profile DPA approach is doing. For future work, it would be possible to consider other strategies for specifying the forcing profile aloft. E.g., instead of blending toward a constant value, it would be straightforward to blend towards a specified geostrophic wind. However, I believe this approach would have challenges of its own because in reality, a well-defined constant geostrophic wind is rarely observed.

14. **Section 4.3**: Why was the power law used when it is well-known to deviate in conditions when stratification is present?

- The power law was used in the absence of another viable wind-speed model. Scattered data interpolation approaches were evaluated (e.g., natural neighbor as used in Allaerts et al, 2023) but this was found to be sensitive to the spatiotemporal distribution of the data. Other approaches add additional uncertainty and uncertainty to the reconstruction process and I felt that the power law was more accessible to wind energy researchers.
- A comment about using natural neighbor as an alternative has been added in Appendix C (line 609–610).
- The fact that the power law is not applicable with stratification is motivation for the partial profile approach, so that the invalid portion of the power-law fit would be ignored. Applying DPA to the full profile was a test to see if the method would "just work", considering that we are mostly interested in wind conditions near the surface; applying IPA was to see if the weak forcing was robust enough to tolerate this inaccuracy aloft.

15. **Figures 7 and 8**: I don't quite follow how the relatively large errors in Figure 7 translate to relatively low errors (*"mean wind speed and direction trends from experiment are generally reproduced."*) in Figure 8. Could you explain this further? For example, from the single-level assimilation approach, I was expecting to see errors in the hub-height wind speed based on Figure 7 and associated discussion, but the spread between DA approaches in Figure 8(a) appears to be very small.

   - If you consider the profiles at hub height (where the winds are plotted in Fig 8), you will find that there is general agreement between methods in Fig 7.

16. **Figure 8**: Can this figure be made more clear as to which 'measurements' are assimilated into each approach (interpolation), and which measurements are not (extrapolation/prediction).

   - Figure 8 has been updated and the different usage of measurement data has been mentioned in lines 345–346: "A distinction has been made in the reference measurements between what was included in the assimilation data (input) and what was used for validation."

17. **Line 314**: What do you make of the friction velocity being predicted well even with the guessed value of z_0 and the fact that the surface heat flux is prescribed?

   - Please note the role of roughness height that I have clarified in the new paragraph at 278–282, "There exists a caveat with regard to the surface BC…"
   - I have added additional comments in the results section (lines 360–363): "For most of the case day, the DPA results track the observed friction velocity (u_*), confirming the relationship between shear stress (described by u_*) and the wind shear in the profile that is exactly matched by DPA. Deviations from observations are seen in the single-level and IPA results, which do not exactly match the observed wind shear and may be sensitive to the choice of roughness height (see discussion in Sect. 4.3)."

18. **Appendix A**: More detail on the actuator disk model (ADM) are required. Numerical implementation. Thrust and power coefficients. Power-yaw relationship, etc.

   - Thank you for point this out. Additional ADM details have been added to lines 575–580.

19. Does FLORIS take inputs from the measurements or from LES?

   - "These LES wind conditions were also used as inputs to FLORIS" is now clearly stated at line 379.

20. **Figure 10**: Why is the impact of yaw misalignment on T2 power so small in FLORIS?

   - The magnitude of yaw misalignment effects is a calibrated parameter in FLORIS — calibrated based on actuator disk LES, actually — that was not varied in this study.
   - The largest decrease in T2 is about 10 % occurring around 0840 UTC when the yaw offset is about 16°. Because the power loss due to yaw misalignment is proportional to

$\cos^{\gamma}(\Psi)$ where $\gamma \approx 2$. This corresponds to a power loss of 0.92, which is close to FLORIS (and the LES) is predicting. The variability in the measured T2 power may be due to wind direction variability and/or actual yaw offset uncertainty.

21. **Line 363**: Turbulence intensity: the discussion presumes that TI estimates from LiDAR are true, but there is a high degree of uncertainty estimating TI from the variance of averaged LiDAR measurements. Worth mentioning.

    - Thank you for the suggestion. I have updated the first discussion paragraph at lines 422–429.

22. **Line 374**: Numerical oscillations from turbine forcing implementations in LES can also be addressed through improved ADM numerical methods [1]

    - Thank you for this reference, I have added a comment at lines 443–444.

23. **Line 383**: Where exactly did FLORIS overpredict the effect of wake steering in Figure 10? I am not sure I see an effect that is significant (statistically?). If you mean at 830, FLORIS seems to already disagree with LES at ~807, so it's not clear what is from wake steering misprediction vs. dynamic errors.

    - You are correct that I am referring to ~830 UTC. The LES predicts partially waked conditions at T3 (Fig 11b) while FLORIS predicts nearly unwaked conditions. I have added "most pronounced around 0828 UTC" at line 451.
    - You have raised a fair point about whether the error is due to wake steering misprediction or dynamic inflow. However, I would say that at ~807, FLORIS and LES are actually in reasonable agreement. T2 has 0° yaw offset so there is no wake steering occurring, and T3 is partially waked (Fig 11a). The partial waking arises from wind direction misalignment, i.e., the wind vector not pointing directly from T2 to T3. Differences in predicted T3 power at this time may arise, e.g., from underpredicted wake expansion or lack of wake meandering in FLORIS. The FLORIS development team has conveyed to me their expectations, that FLORIS "accurately predict trends over the lifetime of a wind project… [and they] do not necessarily expect agreement with instanteous or 10 min averaged conditions" (Appendix A3).
    - I have also added a new paragraph discussing the results during the final set of wake steering periods from ~1030 UTC which I believe exhibits the uncertainty to which you refer. Please see lines 413–417. A comment has also been added in the conclusions: "Even with high-fidelity inflow driven by local measurements, uncertainty still exists due to the dynamic variability in the inflow. This makes it challenging to disentangle the effects of wake steering and from dynamic inflow effects during borderline conditions when a downstream turbine is very nearly or slightly waked." (lines 503–506).

24. After finishing reading the paper, I don't think I agree with this statement in the Abstract *"demonstrating agreement with measurements under partially and nearly waked conditions."* It would be better if you reported the error associated with the background wind conditions (freestream), and the error associated with wake steering predictions (wake effects).

- The other referees also raised a similar concern. Please see the general response at the beginning of this document.