Response to Reviewers for:

An in-depth observational and modeling analysis to explore long-range offshore wakes under different stability regimes

Elizabeth J. McCabe, Jeffrey M. Freedman, Philippe Beaucage, Nick Robinson, Akila Gothandarman

Submitted to *Wind Energy Science*

We would like to thank all the reviewers for their detailed and thoughtful comments. Suggestion and edits have been incorporated into the manuscript and we have responded to all the comments below in red text.

Thank you.

RC1:

The manuscript entitled "An in-depth observational and modeling analysis to explore long-range offshore wakes under different stability regimes" presents a comparative study between aircraft-based measurement data—collected during the WIPAFF project—and WRF simulations for the corresponding period. The manuscript is well-structured and systematically addresses the research questions posed at the outset.

However, while the logical flow is sound, the manuscript would benefit significantly from a more robust discussion of the findings. Specifically, there is a lack of critical interpretation regarding the comparison of wake characteristics between the empirical measurements and the numerical simulations.

Recommendations:

- **Model Reproducibility:** While the WRF model setup is introduced in Section 3.1, a more comprehensive description of the configuration (e.g., specific parameterizations, domain nesting, and boundary conditions) is required to ensure that the results are reproducible by the broader research community.

  Thank you for your comment. Table 1 contains all of the specific parameterization choices I used for my model setup. I used a one-way nested 3-domain setup, with grid resolutions of 9 km, 3 km, and 1 km. I also note that I initialize my model with the ERA5 reanalysis (lines 130-140). I added in Table 1 that the ERA5 is 3-hour temporal resolution and 31 km spatial resolution.

- **Data Methodology:** The acquisition and post-processing of the airborne measurement data are currently described with limited detail. It remains unclear, for instance, what filtering criteria were applied to the raw data or how potential outliers were handled.

  The post processing we did on the aircraft data is described in Section 2.1 (lines 85-105). We did very little post processing of the data. The beginning and end of each flight was removed, as we did not want to consider the times when the aircraft was taking off or landing (added to text on line 141), and we averaged the data into 15sec intervals.

- **Scientific Contribution:** The inclusion of a dedicated **Discussion** section is highly recommended. This would allow the authors to contextualize their findings, highlight the most significant results, and explicitly state the contribution of this analysis to the fields of wind energy and mesoscale modeling.

  Thank you for this suggestion. As you'll now see in the text, I have included a dedicated discussion section. Please see new discussion in lines 480-540.

Detailed Comments

Line 22: Please clarify the term **"turbines facilities"**. It is recommended to use standard industry terminology, such as "wind energy converters" or "wind turbines."

Thank you, this has been corrected to “Wind Turbines”, see line 23.

Line 43: The current sentence structure, particularly the dependent clause in German, is difficult to follow. Please restructure this for improved readability.

I apologize for the confusion; Horns Rev 1 is the name of the wind farm. I re-structured the sentence so it should read more clearly (line 43): “The North Sea offers a long duration of offshore wind deployment, beginning in 2001 with construction of Horns Rev 1, the first modern large-scale (160 MW) wind farm”.

Line 46: The authors should acknowledge the role of Lidar measurements in this context. Several studies have utilized such data; for example, *Cañadillas et al. (2022)* investigated offshore wind farm cluster wakes using long-range scanning wind Lidar and mesoscale modeling.

I have included a reference to LiDAR data here, citing the work done in Canadillas et al. 2022. Please see line 49.

Line 69: Consider referencing *Berge et al. (2024)* regarding their comparison with X-Wakes measurement flights not WIPAFF.

Thank you, this line has been rephrased to cite X-Wakes instead. Please see lines 72-73.

Figure 1a: Please provide a justification for the selection of this specific vertical profile over others. There appear to be significant spatial and temporal discrepancies between the available profiles that should be addressed.

The other flight profiles are taken close to the time of takeoff or landing. We specifically selected profiles that are in open air mid-flight as to be a true offshore wind profile. That is the reason for the profile chosen. I added a note specifying that these profiles are specifically chosen to be away from land. In all the flights, a mid-flight profile was often chosen, as generally occurred furthest from land. Please see added note in line 106.

Line 219: The authors state that wakes from all 41 flights were compared; however, 16 of these are not wake measurements (line 94). Please clarify the methodology used to include these non-wake cases in the comparison.

Thank you. I compared the results from the WRF model to the WIPAFF cases across all 41 events, calculating the phase error and stability for all 41 cases. I agree that the text here is misleading. I have revised the text on line 225 to read: “To validate the WRF model’s ability to properly represent stability regime and long-range wakes in an offshore environment, the WRF model is compared against the 41 WIPAFF flight cases.”

Line 226: *Siedersleben et al. (2018)* already compared WIPAFF data to the WRF model for a single flight rather than the entire ensemble. So this sentence should be rephrased to accurately reflect the scope of that study.

Thank you, I have rephrased this sentence appropriately. Please see lines 234-236.

Table 4: * Regarding the phase error: Were comparisons conducted only at full-hour increments, or did the error always result in full hours? Would analysis at smaller temporal increments yield more precise results?

Comparisons were conducted only at full-hour increments. Although the WRF models were run at 10-min temporal resolution, comparing the time shift between 10-min intervals and hourly did not change the results significantly. Yes, higher temporal comparison could improve the results, but it is difficult to compare at scale and that method was not a good fit for this study. I clarified this in the text on line 244, specifying that the comparison is done at hourly intervals.

How were flights with mean altitudes significantly above hub height (e.g., 180 m +) analyzed for wakes and subsequent time shifts? Please elaborate on this process.

While the examples we focus on in the text, 5/27/17 and 9/6/16 are examples of long-range wakes, as discussed, the method used to determine the offset was based fully on statistical metrics. This is explained in lines 256-260, and below: "To account for a phase error in the analysis, various WRF forecast hours are compared with the flight observations using metrics including the coefficient of determination (r^2), root mean square error (RMSE), and the mean bias error (MBE). The hour offset with the best combination of a high r^2, low RMSE, and low MBE is determined to be the WRF phase error for that case."

As a result, cases without the presence of wakes (e.g., those flown above hub height), were still analyzed for time shift.

Line 325: Please provide the relevant citations for the studies mentioned here.

Thank you, please see added citations on line 335, including Siedersleben et al. 2018; Cañadillas et al. 2020; Ali et al. 2023; and Hahmann et al. 2025.

Line 337: Have other vertical aircraft profiles been examined? The spatio-temporal bias noted may vary significantly depending on the time of day and the geographical position of the measurement.

The flights took profiles at varying times during the day (sometimes morning or afternoon - depending on when the aircraft flew). The location of the profile also varied depending on the wind direction. While I discuss only a few of the aircraft profiles in detail in the text, during the analysis we examined the results in all flights. The data from the aircraft profile is how the stability in Table 1 was determined.

Of course, given that we are working with observational data, our study works within the limitations of our dataset. Looking at additional aircraft profiles was outside of the scope of this study.

Line 363: It is suggested to look into the X-Wakes dataset 8 https://doi.pangaea.de/10.1594/PANGAEA.955382), as it may contain extreme stability cases that could strengthen the analysis.

Thank you for this suggestion, we are aware of X-wakes, but the focus of this study was on the flights during the WIPAFF campaign. The inclusion of these cases is outside of our scope here but will be analyzed further in future work.

Figure 10: Please provide a potential physical explanation for the three outliers in subfigures (a) and (b), where the measured Capped Inversion Height (CIH) is more than triple the WRF-simulated CIH. Were there specific meteorological conditions during these flights that caused this divergence?

Given that a number of these days were near-neutral cases, there was not always significant gradients in specific humidity or potential temperature within the lowest 1000m, so the strongest gradient is somewhat small, potentially leading to larger differences between the WRF and flight results. Also, the flight data is a moving profile (moving both horizontally and vertically), so we are working within the limitations of an imperfect observational dataset. It is expected that there will be outliers. Overall, the relationships shown here between the flight capping inversion height and the WRF CIH and PBL height are remarkably strong. I have added a sentence to the text, line 389-390 noting this.

Line 390: Typo correction: "can be used."

Thank you, this typo has been corrected.

Line 418: Grammar correction: "...profiles **show** that..." (or "...profile shows that...").

Thank you – text has been corrected to “profiles show that”.

Figure 13: The text and labels in this figure are currently too small. Please increase the font size to ensure legibility.

Thank you. I have updated this figure, now Figure 12 (line 440) with bigger labels so it is easier to read. Also, given the fact that the WRF results matched the observations, we decided they didn’t contribute enough to the text. Therefore, the panels in the figure now only show the observational wind roses. Therefore, the overall figure size can also be increased.

Line 460: What is the hypothesized reason for the shift in the theta_v gradient observed around 2012/2013? Furthermore, do the subsequent years show an ongoing trend?

This is an interesting question and is something we looked into significantly. At first, we expected that it must have been due to a measurement error by the FINO3 tower. Which, although still possible, after examining various variables we have no reason to believe there was poor data. Data recovery for the years and various seasons was strong.
This is something that is worth looking into further in future work. It is difficult to make a strong conclusion with limited observational data (only 4 years of unwaked data at the FINO3 tower).
We didn’t analyze data after 2013 given the construction of nearby wind farms. A longer analysis, considering more years should be considered in future work.

Line 489: The discussion comparing the measured and modeled (WRF) wakes requires significant expansion. While these results appear in Table 4, they are underrepresented in the text. Given that this is a core finding, more emphasis should be placed on interpreting the model's performance in both the near- and far-wake regions.

Thank you for pointing this out. Based on your comments and those of other reviewers, I have added a dedicated discussion section to discuss the results from the study in more depth. Please see added discussion, especially regarding the phase error on lines 480-495.

Line 503: Please justify the selection of 2012 for this analysis instead of the years during which the WIPAFF flights were conducted.

Following the analysis we did on the climatological characteristics of the North Sea from 2010-2013, full year WRF simulations had already been run and were available to do this analysis.

RC2:

Review of manuscript wes-2026-55

This manuscript investigates offshore wind farms wakes under different atmospheric stability conditions. The authors compare in situ measurements with model results from WRF.

Overall, the manuscript is clear and well-written. However, there are a few things that could improve the work. Specifically, the figures should be made clearer and the font must be made readable.

Specific comments

Specific humidity, as shown in Figures 6 and 11, is quite different for observations and model results. Please explain/discuss this further.

In Figure 6, the seemingly large differences between specific humidity in the observations and the WRF model result from two things. Firstly, the y-axis is very tight, we are looking at a range of only 1 g/kg here. Secondly, the flight measured specific humidity varies across the wake, and the measurements for specific humidity are higher on the east side of the wake. Please see discussion in the text, lines 393-326: "Air pressure and specific humidity (although slightly higher in the observations on south-east side of the wake) show strong agreement between the observations and model. The small cross-wake differences in specific humidity are expected because the aircraft samples continuously along a moving flight path, whereas WRF represents grid-cell-averaged conditions."

The differences in Figure 11 are again for the same reason. We're comparing a moving flight profile (that varies both vertically and horizontally) to a WRF grid cell at the same location. The resolution in the WRF model is more course, and will not pick up the exact same variations. I've added discussion in this section as well, please see lines 401-410.

The results in part II seems a bit apart from the rest of the manuscript. Consider if it would make more sense to put it into an appendix.

Thank you for this suggestion. I understand that this is a bit different, but we feel that the results from part II are very important to adding context to understanding the North Sea, and how often long-range wakes will realistically occur. We believe the results in part II add a significant contribution to the field, and is work that has been under-represented in previous research. A better understanding of the overall climatological characteristics of the North Sea can also help make this study, and the results, more applicable to other regions, i.e., the New York Bight.

Point 3 in Conclusions seems to introduce results not presented elsewhere. Please move this to an appropriate place in the manuscript (including Fig. 17).

Thank you. You make a good point here. Given your comments and those of other reviewers, I have reorganized the latter half of this manuscript into a discussions and conclusions section. Therefore, the results around Fig. 17 are presented in the discussion and are more relevant to the section and findings.

Technical corrections

Line 22: Is “un-waked” an established term?

I have corrected all instances of “un-waked” to unwaked, following what was used in Houck 2021 (https://doi.org/10.1002/we.2668).

Line 53: “Phase error” is used frequently in the manuscript. Perhaps it would be clearer to specify, at least the first time, that it refers to a time shift?

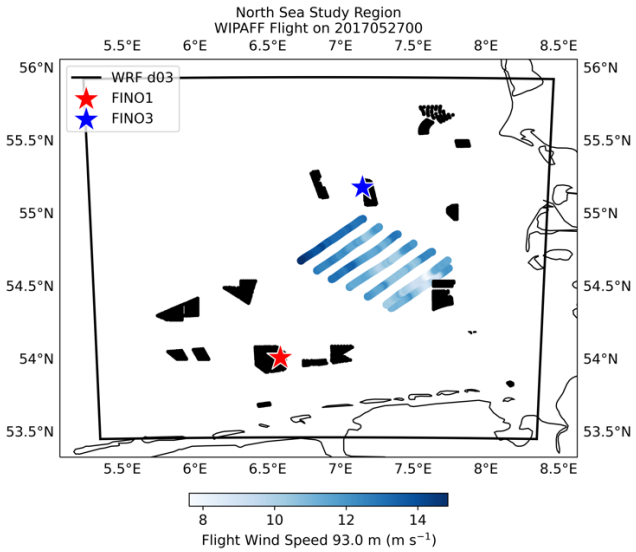
Thank you, I clarified this at the first use of the term “phase shift” on line 53. See added text: Previous analysis on the WIPAFF measurements by Cañadillas et al. (2020) and Berge et al. (2024) note a phase error (i.e., a time shift between the modeled and observed evolution) when comparing these observations to mesoscale models, given that each flight occurs over a period of a few hours.

Line 68: I guess “wake loses” should be “wake losses”?

Thank you, this has been corrected. See line 70.

Figure 2: It is very difficult to find the blue star of FINO3 against the background of black dots.

I have increased the size of the markers for FINO1 and FINO3 and added an outline so they should now be more legible. See below and Figure 2 in text.



Line 155-156: U and V are referred to as gradients. Should it be ΔU and ΔV?

Yes, the U and V have been corrected to ΔU and ΔV. See line 165.

Figure 3: Please check if you can choose another color for “hour offset 0” as it is very difficult to separate from “hour offset 1” in a printed document.

Thank you, this is a good suggestion. It is difficult to pick the color scheme with so many lines. I changed the red line to be gray. See new figure below and updated in Figure 3.

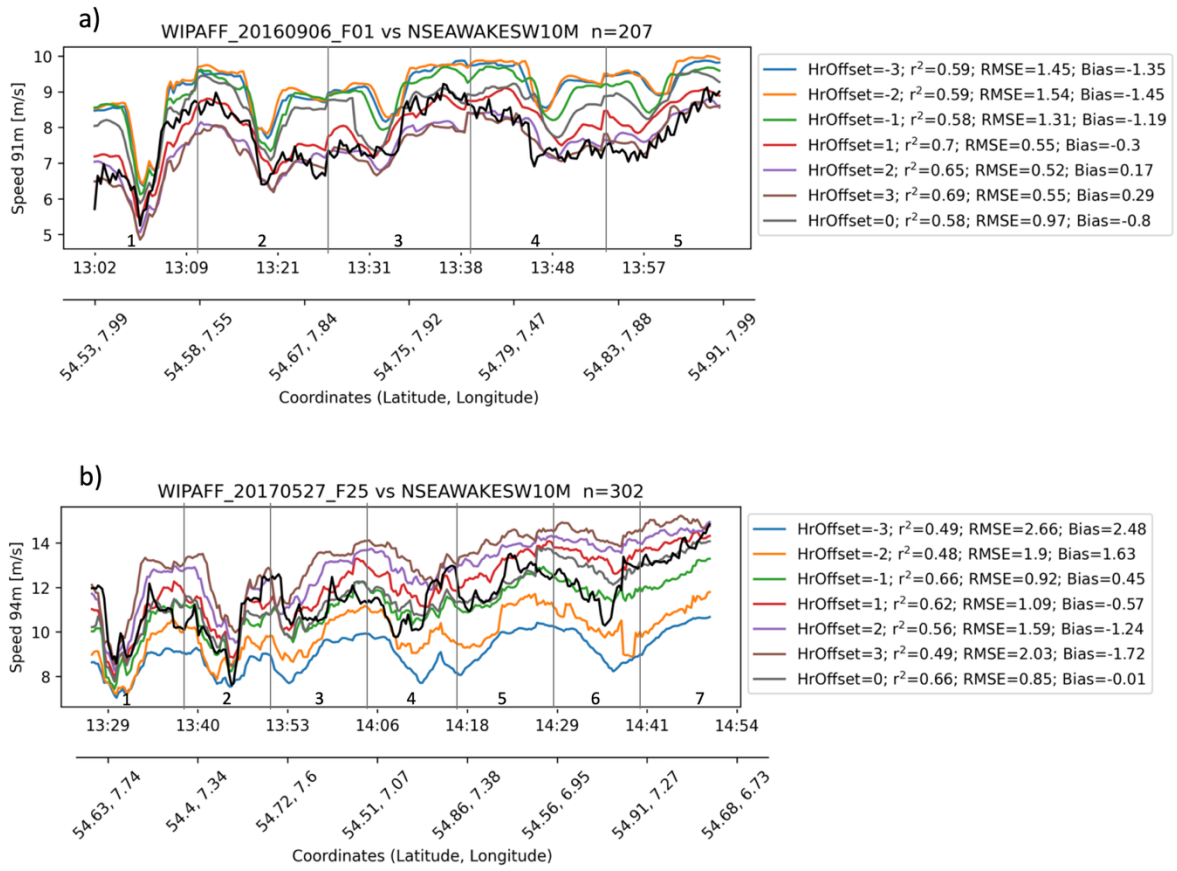


Figure 3: Please add what the grey lines represent (like you do in Fig. 6)

I’ve included a description for the grey lines, see updated figure caption for Figure 3, line 255.

Figures 4 and 5: Font for coordinates is too small to read.

Thank you, this has been corrected. The figures have both been updated to increase the font size. See updated figures in text.

Line 272: Here you write “time synching problem”. Is that the same as the “phase error”?

Yes, I changed “time synching problem” to “phase error” for consistency.

Figure 7: It is difficult to distinguish the “Stability measurement location” from the black dots of the wind turbines.

Thank you. I’ve updated figure 7 to change the scatter point marking the stability measurement from black to orange. See line 325.

Line 376: Please remove extra “that”.

Thank you, extra that has been removed.

Line 418: “hows” -> “shows”

This typo has been corrected.

Figure 13: Font is too small to be readable.

Thank you. I have updated this figure, now Figure 12 (line 440) with bigger labels so it is easier to read. Also, given the fact that the WRF results matched the observations, we decided they didn’t contribute enough to the text. Therefore, the panels in the figure now only show the observational wind roses. Therefore, the overall figure size can also be increased.

RC3:

The manuscript investigates long-range offshore wind-farm wakes and how their behavior changes under different atmospheric stability regimes in the German Bight. The study combines observations and modeling by using airborne measurements collected between 2016 and 2017 over offshore wind-farm wake regions during the WIPAFF project and FINO3 met-mast data to validate WRF model simulations with Fitch wind farm parametrization.

The main scientific focus is on the influence of atmospheric stability, as characterized by the Bulk Richardson number and the atmospheric boundary layer height, on offshore wind-farm wake length and downstream wind-speed deficits, with additional validation of the capping inversion height.

The study finds that the Bulk Richardson number derived from WRF is a reliable metric for classifying offshore atmospheric stability, particularly when using the Businger-Hicks formulation. The paper also suggests that stable offshore conditions can lead to wakes longer than 50 km, and it finds good agreement between WRF-derived capping inversion heights and aircraft profiles.

This is an interesting topic and is relevant for wind-energy applications. However, there are a few limitations that should be addressed. I hope the following comments and suggestions will help improve the manuscript and make it more suitable for publication.

General comments:

- The manuscript would be easier to follow with a clearer structure. The Results and Discussion sections should be more clearly separated before the Conclusions. At the moment, some discussion points are included in Sections 4 and 5, which mainly present the results. A separate Discussion section would help clarify the interpretation of the findings, the study limitations, and the broader implications for offshore wind-energy applications.

  Thank you for this suggestion. I have added a dedicated discussion section to more clearly discuss the results, along with a separate conclusion. Please see lines 480-565.

- It would be great if the study's main contribution and novelty were clearly stated.

  The studies contribution to the field has been added at the end of the introduction, on lines 75-80, and below:

  This study provides the first in-depth analysis validating the WRF model derived bulk Richardson number and capping inversion height to observations in offshore regions. The work done here is increasingly valuable, given the continued development of offshore wind worldwide. The findings in this study can be applied to other regions, and help to establish a robust climatology of stability, and better understand conditions during which long-range wakes are most likely to occur.

- The connection between the validation parts and the climatology could be clearer. Why was the period 2010–2013 chosen for the climatology study?

Thank you for this comment. As mentioned in section 2.2, we used the FINO3 tower data for observational measurements of the offshore environment. As the goal was to understand the overall climatology of the North Sea, and how well the WRF model can represent it, we wanted to study years that were not impacted by wind farm wakes. FINO3 became operational in 2009, but after 2013, became waked by operational wind farms Sandbank and DanTysk. For this reason, we chose to focus the climatology analysis the years between 2010 and 2013, as the tower was not yet surrounded by wind farms.

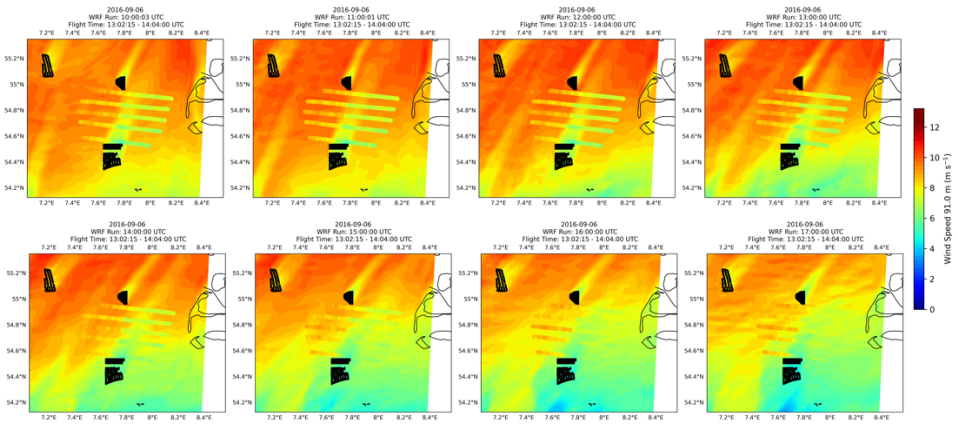
Please see discussion in Section 2.2 (lines 118-119) and at the start of the climatology section (lines 421-429).

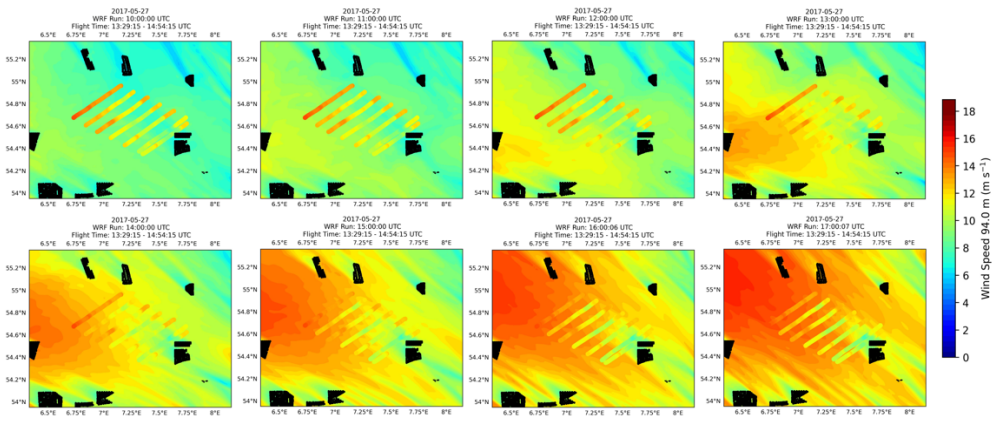
- Not clear whether the validation of FINO3 data includes the cases where there is an impact of wakes from Sandbank or DanTysk.

FINO3 was not impacted from wakes by Sandbank or DanTysk until after 2013, see lines 421-429. For the climatological analysis, FINO3 was not impacted. During the years concurrent with the WIPAFF flights, and the case studies analyzed in 2016 and 2017, yes, the FINO3 tower would have been impacted by nearby wind farms. That is a limitation of the dataset.

- Some figures (for example, Fig 4, 5) could benefit from a colorbar that better highlights details between the observation and the model.

Thank you for your comment. This color bar was difficult to set, given that we are overlaying similar wind speeds. We tested a number of different color bar options, and in order to also ensure we are using an accessible color map, this was the best representation. Please see a previous iteration of figures 4 and 5 below.





- Please add the correct reference to the WIPAFF data throughout the manuscript : Bärfuss, Konrad; Hankers, Rudolf; Bitter, Mark; Feuerle, Thomas; Schulz, Helmut; Rausch, Thomas; Platis, Andreas; Bange, Jens; Lampert, Astrid (2019): In-situ airborne measurements of atmospheric and sea surface parameters related to offshore wind parks in the German Bight [dataset publication series]. PANGAEA. https://doi.org/10.1594/PANGAEA.902845

  Thank you for pointing this out, this reference has been added throughout the manuscript.

Specific comments:

- Line 50; 80: please add the reference Bärfuss et al., 2019

  Thank you, this reference has been been added.

- Line 111—114: Depending on wind direction, measurements at FINO3 may be affected by wakes from the Sandbank and DanTysk wind farms. This means those measurements are not under fully free-stream conditions. Are those cases excluded from the analysis from FINO3? Please clarify.

  As mentioned in my response to your comment above, yes, the FINO3 data that is concurrent with the WIPAFF cases (2016 and 2017) was impacted by the nearby wind farms. However, the reason for choosing years 2010-2013 in Part II, was to pick years where FINO3 did have fully free-stream conditions. This is acknowledged in the text on lines 118-120.

- Line 145-148: Missing ‘)’ and sentence not clear; please check or reformulate.

  Thank you, I re-wrote this sentence, see below and in text on line 152.

  Different methods for estimating atmospheric stability from WRF are evaluated for this analysis. The Monin–Obukhov stability parameter, defined as the ratio of measurement height (z) to the Obukhov length (L), is commonly used to classify atmospheric stability (Monin and Obukhov 1954; Barthelmie 1999; Floors et al. 2011; Archer et al. 2016; Cantero et al. 2022; Rosencrans et al. 2024). However, because the observational dataset does not include in situ surface flux measurements, L cannot be derived from the observations and therefore this

approach cannot be used for the observational analysis. We therefore consider stability metrics that allow for direct comparison between the stability derived from WRF and observations.

- Line 256-261: Would the outcome of the study be different if the time phase difference were not taken into account?

  Yes, however, as shown, the WRF model and observations are clearly offset. This is a known problem with the WRF model. Correcting for the time offset in the model allows us to better compare directly to the observations, and makes the results of this study more robust, while also acknowledging a true limitation of mesoscale models.

- Line 280-290: Could the cold bias in the WRF model influence the stability calculation?

  Of course, but this is a constraint of the study. Correcting for the bias was outside of the scope of this study, but would be interesting to consider in future work.

- Line 320: “before and after” : could it be clearer (for e.g., upwind and downwind?)

  Good suggestion, “before and after” has been changed to “upwind and downwind” see caption for figure 7, line 325.

- Line 360 (profile): missing “s”?

  Thank you, corrected to “profiles”.

- Line 376: “Using this method…”: Please remove the second “that”.

  Thank you, this typo has been corrected.

- In Figure 9, the agreement between WRF and FINO3 appears stronger than the agreement between WRF and the flight profiles, particularly for the lapse-rate metric. This difference should be discussed more clearly, as it may reflect differences in sampling strategy, measurement height ranges, temporal averaging, or the spatial representativeness of the aircraft profiles compared with the fixed FINO3 tower measurements.

  You make a great point here. It is not surprising that the measurements from the FINO3 tower would agree better with WRF, as we are comparing a stationary measurement (FINO3) to one that varies across time and space (Flight profile). Please see added discussion on lines 506-510.

- Line 402–404 and Figure 12: The distribution of points is strongly concentrated around neutral to stable conditions, with relatively few unstable cases. This uneven sampling may partly explain why the relationship between wake length and Rib is unclear. Platis et al. (2020, 2022 suggested that wake length is strongly linked to thermal stability. Could the authors discuss whether a thermal stability metric, rather than Rib alone, may better explain the relationship between wake length and stability? Ref: Platis, Andreas, Marie Hundhausen, Astrid

Lampert, Stefan Emeis und Jens Bange. 2022. 'The Role of Atmospheric Stability and Turbulence in Offshore Wind-Farm Wakes in the German Bight'. Boundary-Layer Meteorology 182 (3): 441–69. https://doi.org/10.1007/s10546-021-00668-4.

I couldn't do this because of limitations in the flight measurements. Although the aircraft measured SST, the next lowest air temperature measurement was at the flight's location, which was approximately 50 m ASL. An air temperature measurement at 50m is too high to give us the ability to use the Air-SST difference as a thermodynamic stability measurement.

- Section 5: The term "climatology" is too strong for an analysis based on four years of WRF simulations and FINO3 observations. The 2010–2013 period provides useful information on multi-year variability. However, it is not long enough to support robust climatological conclusions. Please explain the choice of the period 2010—2013.

  I understand your concern about using the term climatology. I've corrected all instances of "climatology" to either climatological characteristics or analysis.

  As mentioned in an earlier comment, the reason for the selection of the 4-yr period from 2010 to 2013 was due to data availability, as FINO3 was installed in 2009 and became affected by nearby wind farms after 2013.

- Line 475: Is it Fig. 16 or Fig. 14? Please check.

  It is Figure 14 (now, Figure 13). Figure 13 shows the seasonal breakdown of stability and wind speed.

- Figure 14: The results presented are only for 2011 and 2012. However, the climatology section discusses the full 2010–2013 period. It would be helpful to show all four years or provide quantitative evidence that 2011 and 2012 represent the full period.

  Figure 14 has been changed to Figure 13. See note on line 540: "Note the annual stability patterns in 2010 resemble 2011 with 2012 like 2013; therefore, only the years 2011 and 2012 are shown in Fig. 13." As said here, the results in 2010 and 2011 are consistent, and same with the results in 2012 and 2013. There is no additional value by adding the other years to this figure. The goal of this figure is to show the apparent shift in stability regime.

- Line 483—484: Studies based on Satellite SAR also showed that atmospheric wakes can extend beyond 50 km for stable stratification. Please add relevant references here. Please add relevant references here (Christiansen and Hasager, 2006; Djath et al., 2018, 2022…)

  Thank you, relevant citations have been added. Please also see mention of SAR satellite imagery in line 46-47.

- Line 502—505: Please clarify the criteria for selecting 22 cases from the 2012 WRF simulation. This part of the analysis is not limited by airborne measurements. Could the number of cases be increased to achieve a better, more robust outcome?

As noted on line 528, the goal here was to select a better mix of stability regimes, including some very stable and very unstable cases. A subset of 22 included a mix of stable, unstable, and neutral cases. The results here show some relationship between stability and wake length. Of course, more cases would lead to a more robust outcome, but that is outside of the scope of this paper and may be considered more in depth in future work.

- Line 506—508: These statements are important and should be supported by relevant references and further discussion. Please provide some references and further discussion.

  I have moved these results into the discussion section and added and extended discussion on lines 530-540.