

Authors' Response to Reviews of

Assessing the Accuracy of a Three-Year High-Resolution Mesoscale Wind Farm Wake Simulation with Lidar and Satellite Radar Data

Alexandros Palatos-Plexidas on behalf of all authors
Wind Energy Science (WES) Journal,

RC: Reviewers' Comment, AR: Authors' Response, □ Manuscript Text

We would like to sincerely thank the Editor and the Reviewers for their careful evaluation of our manuscript and for the constructive comments and suggestions provided. We highly appreciate the time and effort invested in reviewing our work.

The feedback of the reviewers has been extremely valuable and helpful in improving the clarity, robustness, and overall quality of the manuscript. We have carefully addressed all comments and revised the manuscript accordingly. Below, we summarize the main changes and improvements implemented in the revised version. Detailed, point-by-point responses to each reviewer's comments are provided in the following sections, where all questions and remarks are addressed.

- The BSB metrics evaluation has been improved, as an incorrect hub-height-interpolated value was accidentally used previously by not accounting for the platform height. More specifically, in the first version of the manuscript, we unintentionally used the lidar-measured heights without adding the BSB platform height of 45 m. This led to an incorrect estimation of both wind speed and wind direction at the intended hub height of 107 m.
- The methodology has been improved to provide additional details regarding the SAR images, as well as other aspects that were requested to be clarified.
- The full three-year-long analysis (Section 3.2.1) now includes two additional figures (Figures 7 and 8), presenting wind speed and wind direction biases under different wind direction sectors, thereby providing a more detailed representation of the bias sources at each location.
- In the full three-year-long analysis (Section 3.2.1), hourly means of the lidar measurements and WRF outputs are now used to provide a fairer comparison against ERA5. In the previous version, instantaneous values were used.
- In Section 3.2.3, where we assess the performance of WRF across different wind speed regimes and under southwest and northeast wind directions, two tables (Tables 4 and 5) have been included, presenting bias and cRMSE metrics for each case.
- Additional figures in Section 3.2.3 (Figures 14C and 14D) have been included, presenting wind speed deficits across different wind speed regimes and atmospheric stability conditions.
- In Section 3.2.4, Figure 15 has been updated, and Table 6 has been added, describing wake characteristics under different wind speed magnitudes.
- In Section 3.3, Figure 17 has been updated, and Table 8 has been added, presenting mean absolute error (MAE) and absolute errors (AE) for the analyzed SAR events.

- All results sections have been updated accordingly.
- The Abstract, Introduction, and Conclusions have been revised accordingly.
- The Appendix has also been updated, including the equations for all metrics used and a table with details on the modeled wind turbines.

1. Reviewer #1

RC: *The paper employs WRF with the Fitch wind-farm parameterization to simulate three years of operation in the Belgian–Dutch offshore cluster. Results are compared with Lidar data at four sites and with SAR data, classified by stability regime to assess performance under different conditions.*

Overall, the paper is clear, well organized, and technically competent. However, the study’s novelty is not well articulated. The paper currently reads as a case-study application of an existing parameterization rather than addressing a clear research gap. The introduction lists several limitations of Fitch’s parameterization and mentions multiple validation studies, yet the manuscript does not explicitly explain what new insight this paper brings. Clarifying the main research question and the key take-home messages would significantly improve the manuscript.

AR: Dear Reviewer,

We sincerely thank you for your time and for the detailed, insightful comments. As stated below, we have addressed each of these comments, and believe this has significantly improved the quality of the paper. All changes and additions in the revised manuscript are highlighted in blue and enclosed in boxes.

We should also highlight that we have corrected the interpolation at hub-height for the BSB-lidar cases, resulting in a more accurate representation of the metrics throughout the full manuscript. More specifically, in the first version of the manuscript, we unintentionally used the lidar-measured heights without adding the BSB platform height of 45 m. This led to an incorrect estimation of both wind speed and wind direction at the intended hub height of 107 m.

1.1. General Comments

RC: *1. Please clarify the specific research question or gap being addressed. Is the purpose to validate Fitch’s parameterization, or to evaluate WRF’s long-term predictive skill for offshore wakes?*

AR: The purpose of this study is not to validate the Fitch parameterization in isolation, but rather to assess the long-term predictive performance of the full WRF–Fitch modeling system for offshore wake dynamics in an operationally relevant environment. The analyzed Belgian–Dutch cluster is currently one of the largest offshore wind farm clusters, and we consider it crucial to investigate wake dynamics, as according to the 2030 planned commissions, denser wind farm clusters will be deployed in the North Sea Borge et al. [2025]. We therefore evaluate the model’s ability to reproduce observed wake signatures, stability-dependent wake behavior, blockage, and wind speed deficits using a multi-year dataset. Existing mesoscale studies that use the Fitch (or any other WFP) typically span short periods (days to months) and use resolutions ≥ 2 km, whereas offshore operational wake behavior requires year-long statistics and finer-scale representation. Even the existence of year-long studies is usually limited in resolution (≥ 2 km). By conducting a three-year 1-km simulation supported by four lidars, including one inside the cluster, and SAR snapshots, we provide a unique long-term validation of offshore wake simulations. The duration (three years) as well as the model resolution (1 km) allow us to quantify the wakes in both short-climatological terms and at high mesoscale resolution, as there are not many such studies that exist. The Fitch scheme is an integral component of this setup, but its evaluation is embedded within the broader assessment of the ability of the model to simulate wakes over longer periods and provide better statistics.

To reflect this in the revised manuscript, following modifications have been introduced. The abstract has been changed:

The rapid expansion of wind farm installations in the North Sea results in an increased need for understanding their influence on the local atmosphere, as well as the interactions between them. Wind farm operation and power production are affected by wakes produced both within the wind farm and by upwind wind turbines. Accurately estimating these impacts requires robust mesoscale atmospheric modeling capable of capturing extended wind speed deficits and their dependence on atmospheric stability. This study presents a three-year-long mesoscale analysis using the Weather Research and Forecasting (WRF) model at a horizontal resolution of 1 kilometer, a comparatively long and high-resolution configuration within the context of offshore wake research. The simulations are evaluated against four lidars located in the Southern Bight of the North Sea in the vicinity of the 3.7 GW Belgian-Dutch offshore wind farm cluster, providing an extensive observational basis that covers upstream, intra-farm, and downstream flow conditions. Coupling the mesoscale atmospheric model with the Fitch wind farm parameterization (WFP) scheme significantly improves simulation accuracy, particularly in regions frequently affected by wake effects. An evaluation across different atmospheric boundary layer stability regimes shows that the model performs best under less extreme stability, while a detailed assessment of upstream, intra-cluster, and downstream wake characteristics highlights the added value of the Fitch WFP scheme for multi-year offshore applications. Finally, synthetic aperture radar images from selected wake events are compared with the model output, demonstrating that the WFP scheme effectively captures the wake structures at the analyzed timestamps. Together, these results provide one comprehensive long-term evaluation of offshore wake behavior in a dense wind farm cluster, helping to address the current gap in mesoscale wake studies.

We have updated the introduction (lines 63-81) as follows:

The current work presents an assessment of model performance and analysis of wake effects from a three-year-long high-resolution (1 km) WRF simulation of the Southern Bight of the North Sea, with a specific focus on the 3.7 GW Belgian-Dutch wind farm cluster. This cluster consists of several high-density wind farms installed next to each other, making it the largest offshore cluster operational today, hence resulting in a prime test case for studying wind farm wakes. Extending the analysis over a three-year period enables a comprehensive characterization of seasonal patterns and temporal variability. The study is driven by two primary objectives that address notable gaps in current offshore wake research. The purpose of this study is not to validate the Fitch parameterization in isolation, but rather to assess the long-term predictive performance of the full WRF–Fitch modeling system for offshore wake dynamics in an operationally relevant environment. First, we assess the three-year WRF–Fitch simulation performance using lidar measurements positioned both within and around the Belgian-Dutch cluster, enabling an evaluation of model performance under varying ABL stability regimes. This is particularly relevant given that long-term, wake validations at 1 km (assessed under different stability regimes) resolution remain rare, especially over multi-year periods and in densely built offshore environments. The strategic placement of the lidar probes allows us to assess how the model represents inflow conditions, intra-cluster wake dynamics, and downstream recovery for different wind directions and wind speed regimes, and the availability of four lidars spanning upstream, downstream, and intra-farm positions provides a uniquely comprehensive observational basis for this evaluation. Second, we evaluate the performance of WRF for a set of specific wake events by comparing the simulated wake signatures with satellite-retrieved wind fields. This dual comparison, quantitative through lidar transects and qualitative through two-dimensional SAR wake structures, provides a robust, multi-sensor estimation of wake deficits, blockage effects, and model skill. Together, these objectives provide a comprehensive long-term evaluation of offshore wake behavior in a high-density wind farm cluster.

RC: *2. Several metrics show that the wind-farm (WF) simulations do not outperform the no-wind-farm (NWF) simulations, which seems inconsistent with some of the conclusions.*

AR: Results and conclusions have been thoroughly updated in the revised manuscript. In summary, the slightly larger negative bias in the WF simulations can be attributed to the fact that WHi, EPL, and LEG are located close to the Belgian–Dutch wind-farm cluster, where wake effects reduce wind speeds downstream. Consequently, the improved biases in the NWF runs likely result from partial error cancellation with the negatively biased ERA5 input, whereas these errors become slightly amplified in the WF simulations (while still remaining noticeably smaller than in ERA5). This behavior is consistent with previous studies showing improved performance of the Fitch scheme in near-wake regions [García-Santiago et al., 2024, Ali et al., 2023, Fischereit et al., 2024a]. The validation analysis has been expanded accordingly, now including wind-speed and wind-direction bias metrics per wind-direction sector (polar plots in Figures 7 and 8). Additional relevant discussion is provided in our responses to comments 8, 9, and 11.

Furthermore, the transect analysis in Section 3.2.3, Tables 4 (for SW) and 5 (for NE) wind directions, respectively, have been added to assess bias and cRMSE along upstream, intra-farm, and downstream regions. These results show that the WF simulations reduce the bias in intra-farm and downstream areas, particularly for wind speeds between 5 and 8 m s⁻¹ under SW wind direction and between 8 and 11 m s⁻¹ when wind direction is NE. Finally, mean absolute error (MAE) between the model and the 10-meter SAR wind speeds across the transects, as well as the absolute error (AE) relative to the lidars, have been added in Section 3.3 (Table 8). These metrics consistently show that the WF runs outperform the NWF simulations for the analyzed events. An updated portion of the conclusions addressing this point is also included in our response to comment 13.

1.2. Specific Comments

RC: *3. Line 105: Please specify whether one-way or two-way nesting was used. I assume one-way feedback given the use of nested inner domains (d03).*

One-way nesting is used in this study. Section 2.1 has been updated accordingly (lines 96-97):

Furthermore, one-way nesting is applied to prevent any potential noise from the innermost domain from propagating into the parent domains.

RC: *4. Equations (1-3): Fitch’s implementation in WRF applies an energy correction to the momentum and TKE tendencies. The equations presented should reflect the actual WRF formulation, not only the original Fitch et al. (2012) expressions.*

AR: We thank the reviewer for this comment. The energy correction term ensures that the vertically distributed extraction matches the turbine power curve at hub height. We have included the energy correction terms in Equations 1 and 2, denoted by *ec*, in Section 2.1. However, we believe that fully detailing this in the manuscript would distract too much attention from the main storyline. Instead, we add a clarifying sentence to the discussion of Equations 1 and 2 in Section 2.1.1 (lines 132-134) as follows:

The principal equations for describing the influence of the wind farms on the mean flow in a discrete Cartesian computational cell i, j, k are defined as

$$\frac{\partial |V|_{ijk}}{\partial t} = - \frac{N_t^{ij} C_T \left(|V|_{ijk} \right) |V|_{ijk}^2 A_{ijk}}{2(z_{k+1} - z_k)} ec, \quad (1)$$

$$\frac{\partial \text{TKE}_{ijk}}{\partial t} = \frac{N_t^{ij} C_{\text{TKE}} \left(|V|_{ijk} \right) |V|_{ijk}^3 A_{ijk}}{2(z_{k+1} - z_k)} ec, \quad (2)$$

$$\frac{\partial P_{ijk}}{\partial t} = \frac{N_t^{ij} C_P \left(|V|_{ijk} \right) |V|_{ijk}^3 A_{ijk}}{2(z_{k+1} - z_k)}. \quad (3)$$

These equations describe the change in horizontal wind speed $|V|$, the wind power production P , and the turbulence enhancement TKE due to wind turbines, where z_k defines the height at model level k , and A_{ijk} is the cross-sectional rotor area of one wind turbine between the height levels k and $k + 1$. Furthermore, N_t^{ij} is the number of turbines located within cell (i, j) . The thrust coefficient C_T and power coefficient C_P depend on the wind turbine type and are obtained from a mix of public and confidential data. The TKE coefficient is defined as $C_{\text{TKE}} = \alpha(C_T - C_P)$, with $\alpha = 1$ in the original Fitch implementation, driven by the idea that momentum lost by the mean flow, which is not converted to electrical power, is transformed into turbulence. However, it was found that this can lead to a significant overestimation of the TKE compared to high-fidelity simulations [Archer et al., 2020]. Although there remains significant debate over the appropriate value of α [Larsén and Fischereit, 2021, García-Santiago et al., 2024], this study uses the current default value in WRF, i.e., $\alpha = 0.25$. Last, ec is the energy correction factor that ensures that the vertically distributed extraction matches the turbine power curve at hub height. The wind speed deficit due to the wind turbine influence is defined as the normalized difference between the unperturbed and the wind-farm-affected wind speed, i.e., $\Delta U_{\text{wake}} = (U_{\text{NWF}} - U_{\text{WF}}) / U_{\text{NWF}} \times 100$.

The energy correction factor is defined in the WRF source code as:

$$ec = \begin{cases} \frac{P_{\text{hub}}}{P_{\text{blade}}}, & \text{if } P_{\text{hub}} \neq 0 \text{ and } P_{\text{blade}} \neq 0 \\ 1, & \text{otherwise,} \end{cases} \quad (4)$$

where:

$$P_{\text{hub}} = \frac{1}{2} \rho C_P A_{ijk} V_{\text{hub}}^3, \quad (5)$$

$$P_{\text{blade}} = \sum_{\text{bl}} \frac{1}{2} \rho C_P |V_{ijk}|^3 A_{ijk} \Delta X^2. \quad (6)$$

In Equations (5), and (6), $\rho = 1.23 \text{ kg m}^{-3}$ is the air density under standard atmosphere conditions, V_{hub} is the hub height wind speed, and bl is the number of height levels across the area swept by the turbine blades. Turbulence advection is activated in the high-resolution domains.

RC: 5. Line 182: The statement “The wind speed is calculated as a reflected signal” is unclear. Which wind speed do you mean: surface, 10 m, or another height? Note that the 10 m wind speed derived from SAR

is inferred from radar backscatter by assuming neutral stability. So, it is not a measured quantity, but a processed one.

The overall text in Section 2.3.2, has been updated as follows (lines 191-201):

The SAR data used in this study were retrieved by the Sentinel-1 European Space Agency mission. More specifically, there are two satellites, Sentinel-1A (2014-present) and Sentinel-1B (2016-2021), equipped with C-band Synthetic Aperture Radar (SAR) instruments operating at 5.405 GHz, and both satellites follow a sun-synchronous polar orbit at an altitude of 693 km [Hasager et al., 2024]. *Focusing on the area of interest, including the Belgian-Dutch wind farms cluster in the Southern Bight of the North Sea, as well as additional upstream and downstream areas to study potential wake structures, the analyzed time instances occur approximately at 06:00 UTC every twelve days, and at 17:40 UTC every six days, respectively (Sentinel-1A, 2021-2023). The operation of Sentinel-1B during 2021 provided an additional set of overpasses at around 17:32 UTC every six days. This corresponds to approximately 8 to 10 SAR snapshots per month over the area of interest during the analyzed period. The wind speed at 10-meter height is then estimated by applying the CMOD5N geophysical model function [Hersbach et al., 2007], as explained also in the work of Siedersleben et al. [2020]. This indirect retrieval method assumes that radar backscatter from the sea surface, which reflects centimeter-scale waves generated by instantaneous wind stress, can be converted into wind speed at 10-meter height [Ahsbahs et al., 2017, 2018]. The wind fields are provided at a 500-meter pixel resolution through the Global Wind Atlas Science Portal of the Danish Technical University (DTU) Wind Energy Department (<https://science.globalwindatlas.info/#/map/satwinds> last access: 30 July 2025).*

RC: 6. *For completeness, please include a description of the EMD metric in the Appendix.*

AR: We have addressed this in the revised manuscript (Appendix).

The Earth Mover's Distance is assessed using the Wasserstein first distance (WD). WD is defined as a distance between two probability measures P and Q on a metric space $(\mathcal{R} \times \mathcal{R})$ by

$$\text{EMD}(P, Q) = \inf_{\gamma \in \Gamma(P, Q)} \int_{\mathcal{R} \times \mathcal{R}} \|x - y\| d\gamma(x, y), \quad (7)$$

where $\Gamma(P, Q)$ denotes the set of all couplings with marginals P and Q .

RC: 7. *Line 256/Eq. A4: Clarify how wind direction is computed. Are wind components rotated from model to Earth coordinates?*

AR: Wind speed (only horizontal components are used in the study both from the lidars and WRF) and wind direction are computed directly from WRF, using Python. The wind components are indeed rotated to Earth coordinates. It has been included in the manuscript, in Section 3.1, lines 255-256:

All results presented in this study use wind speed and direction rotated to Earth-fixed coordinates.

RC: 8. *Figure 5: Why does including wind-farm effects increase the bias relative to observations at most sites, given that both WF and NWF simulations share the same background bias from ERA5?*

AR: At WHi, EPL, and LEG, bias is already negative in the ERA5 data, which propagates into the WRF simulations and results in a negative bias in the NWF runs as well. The slightly larger negative bias in the WF simulations

arises because WHi, EPL, and LEG are located close to the Belgian–Dutch wind farm cluster, where wake effects generate reduced wind speeds downstream. Consequently, the improved biases in the NWF runs likely stem from error cancellation related to the biased ERA5 input, whereas these errors become slightly amplified (while remaining noticeably lower than the ERA5 errors) in the WF simulations.

This paragraph in Section 3.2.1 (lines 294–301) has been added:

The bias score (Figure 5A) shows that both ERA5 and the NWF simulation exhibit a consistent negative bias at all sites except BSB, a consequence of the inherent ERA5 underestimation that propagates into WRF. The additional wind speed reduction introduced by the Fitch WFP leads to a slightly more negative bias in the WF run, although WRF generally aligns better with the lidar observations than ERA5. **Part of the improved bias in the NWF runs likely results from an error-cancellation effect with the negatively biased ERA5 input, whereas in the WF simulations, these errors become slightly amplified due to wake-induced deficits, but remain lower than those in ERA5.** At BSB, both ERA5 and the NWF runs show a positive bias, while the WF simulation produces a smaller negative bias, corresponding to an improvement of about 82% over NWF and 73% over ERA5 in the near-wake under the predominant southwesterly flow.

In addition, in the analysis of SW, and NE wind direction in Section 3.2.3, in Figure 12, and Table 4, we observe that in the upstream location (WHi under SW wind conditions), there is always a negative bias, most likely propagated from ERA5 and the overall WRF performance. This under specific conditions, for example, in Table 4, under $[3-5) \text{ m s}^{-1}$ wind speeds, BSB provides a negligible bias in the NWF simulations that is higher in the WF. We consider that this can happen due to error cancellation. Similarly, under NE wind conditions in Table 5, when the wind speed is low, the bias at BSB is lower in NWF simulations compared to the WF.

In the conclusions, this paragraph (lines 593–605) has been updated as well:

Our findings demonstrate that the use of the Fitch WFP scheme, in combination with this particular WRF configuration, captures the wake effects and provides an improvement in model performance relative to a simulation without wind farms at the intra-farm region. First, focusing on the time-averaged wind speeds over the full analysis period, the WF simulations yield a substantial improvement in both bias and EMD, with reductions of approximately 82% and 73% compared with the NWF simulations, respectively, in the intra-cluster (BSB location). **Furthermore, when analyzing the wind speed bias under different wind directions, both NWF and WF configurations reduce the negative bias inherited from the ERA5 data, with the results indicating that the Fitch scheme introduces an additional deficit downstream of the Belgian-Dutch cluster.** However, within the intra-farm area, the Fitch scheme provides a more accurate wind speed distribution across all wind directions, underscoring its importance for representing strongly-waked environments. Beyond the improvements in bias and EMD, the WF simulations perform similarly to the NWF runs in terms of cRMSE and correlation, with only small and marginal differences. Similarly, wind direction metrics are also improved in the WF simulations, as bias is reduced by approximately 51% at BSB, while an analysis of wind direction biases by sector reveals that WF can perform better than ERA5 for wind directions from northeast to south.

RC: 9. Figure 5 (cRMSE): ERA5 appears to outperform both WRF setups when compared to centralized error. Please discuss why.

AR: Following the previous comment, Section 3.2.1 has also been updated as follows in lines 316–323:

The larger cRMSE in WRF compared to ERA5 is expected because the coarser ERA5 fields do not resolve kilometer-scale wind speed fluctuations, resulting in a smoother signal and lower instantaneous error. In contrast, the higher spatial and temporal variability represented in WRF leads to phase mismatches between simulated and actual wind fluctuations, producing larger instantaneous discrepancies that manifest as higher cRMSE. The study of [Pronk et al., 2022] shows that ERA5 outperforms WRF simulations in terms of cRMSE and correlation coefficient, both on land and offshore, under different stability conditions. Additionally, a consistent negative bias in wind speed is reported. These findings align with the WRF sensitivity analysis of Olsen et al. [2025] in the North and Baltic Seas, indicating that ERA5 outperforms the various WRF configurations in terms of cRMSE and correlation.

RC: *10. Model sensitivity: Since the WF simulation rarely performs best (only EMD metric at BSB location), could parameter choices (e.g., TKE coefficient = 1 instead of 0.25) influence the results? A brief sensitivity analysis might help.*

AR: Dear reviewer, thank you for your comment. We have conducted a brief sensitivity analysis on the a values on the Fitch WFP scheme, resulting in no significant differences. More precisely, two additional simulations for a period of about 6 months (mid-January to mid-July 2022, since WHi measurements are available during that period) using $a = 0$, and $a = 1$ have been performed. The main conclusions are described below:

- Bias values across the WF simulations are overall very similar. While some differences are observed at individual locations (e.g., slightly lower bias for $a = 1$ at BSB), no single value of a consistently outperforms the others across all sites.
- cRMSE, EMD, and correlation (r) values are likewise very close between the WF simulations, with only minor, site-specific differences and no systematic advantage for a particular value of a .
- Regarding the wind direction metrics, similar trends to the simulations with $a = 0.25$ are indicated in the rest of the WF simulations, while ERA5 results in better metrics as explained in the response to comment 9.

Overall, this assessment indicates that the sensitivity of the wind speed and wind direction metrics analyzed in this study to the choice of a is limited, and that differences between configurations are small. It is also worth noting that the sensitivity assessment was conducted over a selected six-month period and is therefore intended to provide guidance on parameter choice rather than a comprehensive evaluation across the full 2021–2023 analysis period. We also note that the value of $a = 0.25$ was originally introduced to improve the representation of both wind speed and turbulent kinetic energy (TKE) in the Fitch WFP scheme (e.g., Archer et al. [2020], García-Santiago et al. [2024]), which is not (TKE) evaluated in the present study but remains an important consideration. The value $a = 0.25$ was therefore retained as a reasonable and balanced choice

RC: *11. Figure 6: The same issue applies to wind-direction metrics.*

AR: Wind direction metrics are better in the WF simulations in comparison to the NWF runs. However, compared to ERA5, there is indeed a worse matching for the same reasons explained in the previous comment, as well as in the studies of Pronk et al. [2022], and Olsen et al. [2025].

In addition, in the analysis of wind direction under different wind directions, in Figure 8 (in the revised manuscript), it appears that ERA5 does not always perform better when wind direction is from 0 to 180°

(clockwise) (e.g., at WHi, EPL, and LEG WF is better in these cases), but it is slightly more accurate when wind goes from 180 to 360° (clockwise). This is also mentioned in the response to comment 8 (lines 336-340):

Similar to the wind speed bias per wind direction, Figure 8 presents the wind direction bias. Although the mean bias in Figure 6A suggests a closer agreement between the lidars and ERA5, WRF provides a better representation of the wind directions when the wind originates from the north to south-southeast (0 to 180°). Conversely, for westerly wind directions between 180 and 360° ERA5 performs better. At BSB, the mean bias is nearly identical between the WF runs and ERA5, but ERA5 shows a slightly closer agreement with the lidar measurements across all wind directions.

RC: *12. Figure 15: How do you reconcile the comparison with SAR, which assumes neutral stability, while WRF simulates varying stratification? The WRF–Lidar comparison may be more representative, and the SAR transects can be more of a qualitative reference.*

AR: Cases A and C, where WRF indicates stable-neutral and unstable-neutral conditions, respectively, align well with the neutrality assumption in SAR imagery. In contrast, cases B and D, characterized by very unstable stratification at the analyzed locations, introduce uncertainty both in the SAR neutrality assumption and in the stability classification derived from L in WRF. More specifically, Section 3.3. has been updated accordingly: (with bold in lines 544-549 is the part that addresses this specific comment)

Figure 16 presents the comparison of SAR and WRF wind fields at 10-meter height. The left column shows the SAR-derived wind speeds, denoted by the subscript 1. The middle column depicts the WF wind speeds, indicated by a subscript 2, and in the right column, the reference NWF wind speeds are illustrated, and denoted with a subscript 3, as a reference case. In case A, higher wind speeds are also related to stable-neutral conditions and a shallow boundary layer (see Table 7). The wake structure of the Belgian-Dutch wind farm cluster is well captured by the WF simulation. However, a slight deviation in wind direction is observed in WRF, exhibiting a positive bias toward more westerly flow. Similar bias has also been highlighted in the wind roses (Figure 4) and in the wind direction validation heatmap in Figure 6A. The prolonged wake extending more than 50km downstream under stable-neutral stratification agrees with the findings of Cañadillas et al. [2020] and Platis et al. [2018], who also report that wake structures are more pronounced and longer during stable conditions. Furthermore, Siedersleben et al. [2020] examined a stable event simulated with the Fitch scheme in WRF (version 3.8.1), comparing the original formulation ($\alpha = 1$) with a variant excluding the added TKE term, and showed that including the added TKE source under stable stratification results in a more realistic representation of the prolonged wake of a SAR-retrieved 10-meter wind speed pattern. In our study, although we use a finer grid spacing (1km instead of 1.67km) and a reduced TKE coefficient ($\alpha = 0.25$), which limits the influence of the added TKE term, we observe a very good agreement between the SAR image and the WF simulations during stable conditions. Furthermore, Ali et al. [2023] evaluated several WFP schemes to characterize the wake of a specific event using 10 m SAR data. Their results show that, when comparing the Fitch model with $\alpha = 1$ and $\alpha = 0.25$, the latter produces a longer wake downstream of the wind farms, similar to the behavior observed in case A analyzed in the present study.

In case B, the wake is weaker, but still detectable both in SAR image and in the WF simulation. In this case, stability conditions are very unstable, and the PBLH is around 700 meters, explaining the attenuated wake pattern. Case C is characterized by higher wind speeds, L varies from -300 to -450 m, and PBLH is approximately 300 m across the three locations, resulting in a more unstable-neutral stratification profile, resulting in the observed downstream wake of about 40 to 50km. Similar wake

patterns but slightly shorter, about 25 to 35km have been observed during neutral stratification in the study of Platis et al. [2018]. Last, in case D, very unstable ABL conditions prevail, and the PBLH varies between 250 meters upstream at the WHi location to 358 meters downstream at the EPL area. This is the only SAR image with clear wakes for SW wind conditions. Overall, the patterns of the wakes are accurately represented by the WF simulations, highlighting the importance of using the Fitch WFP scheme in WRF. **Although SAR assumes neutral stability to compute the 10-meter wind speed, there are discrepancies with the WRF model, especially in cases B and D, where very unstable stratification dominates based on the values of Obukhov length. Uncertainty exists regarding the assumption of SAR image neutrality used to calculate wind fields, an issue that has also been reported in previous studies across different applications Hasager et al. [2011a], Ahsbaks et al. [2017]. Furthermore, there is uncertainty in the estimation of L within WRF, as it strongly depends on the PBL and surface schemes applied in each configuration.**

In addition, in the revised manuscript, we provide a comparison of WRF against SAR and lidar (when available) for the four analyzed events using mean absolute error (MAE) and absolute error (AE). This has been added in the revised text in Section 2.3.4 (lines 242-245):

Last, for the SAR transects, we compute the mean absolute error, $MAE = \sum_1^n |U_{SAR} - U_{WRF}|$, where n is the number of points across the upstream, intra-farm, and downstream regions of each transect (70 points in total after interpolation of the SAR and WRF data). In addition, for the same time instances, we evaluate the point-wise absolute error, $AE = |U_{lidar} - U_{WRF}|$, using the lidar observations as reference.

The results are presented in Section 3.3. Table 8 has been added and the text has been revised accordingly in lines 554-585:

Therefore, to further evaluate the performance of the WRF model, hub height-lidar measurements are incorporated when available, depicted as gray scatter points in Figure 17. In addition to the qualitative assessment of the wind speeds in Figure 17, Table 8 presents the MAE across the upstream, intra-farm, and downstream transect regions, as well as the point-wise AE, when lidar measurements are available, and the evaluation methodology is described in section 2.3.4. In cases A, B, and C (NE wind flow), the transect follows the direction EPL (upstream) → BSB (intra-farm) → WHi (downstream), whereas case D corresponds to a SW flow event, where the sequence reverses.

Case A shows good agreement in the representation of the transect wake. Specifically, while the WF simulation tends to underestimate wind speeds upstream and within the Borssele zone, it accurately captures the wind speed deficit in the Belgian wind farms and downstream of the wind farm cluster. This is also reflected in Table 8, where only in the upstream region WF generates a slightly higher MAE = 0.49 m s^{-1} , compared to MAE = 0.37 m s^{-1} in the NWF simulation. However, the improvement between the WF against the NWF simulations in MAE in the intra-farm region and downstream is 81% and 85%, respectively. Similarly, in Case B, both the intra-farm and downstream wind speed patterns are well represented by the WF simulation. Although the overall trend is accurately reproduced, the WF simulation tends to overestimate the wind speeds upstream and downstream of the wind farms. Additionally, a flow acceleration is observed as the wind exits the Belgian-Dutch cluster. This pattern is consistent with the broader flow conditions downstream of the Belgian-Dutch cluster, since a similar acceleration is evident in the SAR observations and in both the WF and NWF simulations. Both the lidar measurements and the WF simulation effectively capture the intra-farm wind speed deficit and the downstream flow acceleration. Furthermore, in Table 8, MAE and AE are better in the WF than NWF

simulation across the full transect regions.

Case C also features NE incoming wind conditions. Similar to the previous cases, the WF simulation successfully captures the wind speed deficit trends both within the intra-cluster zone and in the downstream wake recovery region. Both the NWF and WF simulations show noticeable discrepancies upstream at EPL compared to lidar observations with an AE reaching above 3 m s^{-1} . This is significantly improved both at BSB and downstream at WHi, where WF simulations outperform the NWF runs both in terms of MAE across the transect regions and AE against the lidar measurements. During this event, the Obukhov length at EPL is relatively high and negative, indicating unstable to near-neutral conditions. This may be attributed to less negligible shear, which can contribute to increased deviations from the lidar measurements. In the NWF simulation, an unrealistic and nearly linearly increasing pattern in wind speed of approximately 0.5 m s^{-1} between EPL and WHi is observed. Case D is the only selected timestamp with a SW wind direction. Although the SAR data exhibit considerable noise, particularly within the intra-farm region, the WF simulation is still able to reproduce the wind speed deficit and the downstream wake-recovery trend. In this case, only EPL lidar measurements are available, which show good agreement with the WF simulation, with a slightly improved AE compared to the NWF run. Furthermore, the MAE provides a more accurate matching with the 10-meter SAR wind speeds across the three transect regions in the WF simulations.

RC: *13. Line 441 and Conclusions: The claim that “the model is systematically validated” and that “WF simulations improve the general accuracy” is not fully supported by the presented results. Please moderate these statements.*

AR: Statements like the aforementioned one have been moderated. Conclusions have been updated overall, based on updated results and the comments of all three reviewers. For instance, lines 593-605 have been changed accordingly (see response to comment 8 as well).

Our findings demonstrate that the use of the Fitch WFP scheme, in combination with this particular WRF configuration, captures the wake effects and provides an improvement in model performance relative to a simulation without wind farms at the intra-farm region. First, focusing on the time-averaged wind speeds over the full analysis period, the WF simulations yield a substantial improvement in both bias and EMD, with reductions of approximately 82% and 73% compared with the NWF simulations, respectively, in the intra-cluster (BSB location). Furthermore, when analyzing the wind speed bias under different wind directions, both NWF and WF configurations reduce the negative bias inherited from the ERA5 data, with the results indicating that the Fitch scheme introduces an additional deficit downstream of the Belgian-Dutch cluster. However, within the intra-farm area, the Fitch scheme provides a more accurate wind speed distribution across all wind directions, underscoring its importance for representing strongly-waked environments. Beyond the improvements in bias and EMD, the WF simulations perform similarly to the NWF runs in terms of cRMSE and correlation, with only small and marginal differences. Similarly, wind direction metrics are also improved in the WF simulations, as bias is reduced by approximately 51% at BSB, while an analysis of wind direction biases by sector reveals that WF can perform better than ERA5 for wind directions from northeast to south.

In lines 611 to 617, there is an updated version in the conclusions:

Extending the performance metrics to wake characteristics under different wind speed magnitudes, and focusing on SW and NE wind directions, the analysis revealed several consistent patterns. First, under SW wind direction, the WF simulations accurately capture the wake structure when wind speeds range between 5 to 8 m s⁻¹, providing an almost negligible bias at BSB and downstream at EPL. Under NE wind direction, the optimum performance of WF simulations occurs at wind speeds between 8 and 11 m s⁻¹. Although the model tends to overestimate intra-farm wind speed deficits at BSB under higher wind speed conditions, it shows comparable improvements in long-distance downwind wakes at EPL (for SW winds) and WHi (for NE winds), where smaller bias and cRMSE values are obtained.

And in lines 623 to 634:

Furthermore, a comparison with SAR images at specific events shows that WF performs well in representing the wake structure generated by the Belgian-Dutch cluster. Both in the intra-farm and the wake recovery regions, WF provides a realistic wind speed deficit. To assess the performance of the model during these events, MAE and AE are computed. More specifically, MAE is estimated across the upstream, intra-farm, and downstream regions, while AE is calculated at each lidar location when measurements are available. Building on the previously discussed improvements in the qualitative wake-pattern representation, the error metrics further confirm that the WF simulations provide a better alignment with observations than the NWF runs across all locations, with the most pronounced improvements occurring in the intra-farm and downstream regions.

This study can accurately evaluate the model performance by utilizing upwind, intra-farm, and downwind lidar measurements as well as SAR images at specific events. Systematic biases in both wind speed and direction are evident, particularly in the wind speed metrics. However, it should be highlighted that the model is driven by ERA5 hourly data, which consistently introduce biases, affecting the boundary conditions in WRF and propagating wind speed underestimations to the analyzed locations.

RC: *14. The reported 36 % bias reduction appears inconsistent with Fig. 5a, where WF performs worse at most sites. Please clarify how this number was derived.*

AR: That was a typo: we were referring only to the BSB location. However, the entire paragraph in the conclusions has been changed. The detailed paragraph is attached as a response to the previous comment, 13, lines 593-605 in the revised manuscript.

RC: *15. The suggestion that future work should “benchmark different parameterization schemes” is valid but not novel as this topic has been addressed many times in the literature, particularly Fitch’s parameterization and the Explicit Wind-farm Parameterization (EWP).*

AR: It’s a general recommendation for WRF parameterization schemes, not limited to WFP schemes but also including PBL schemes, as they play a key role in turbulence production in WRF and are important for assessing model performance. In this study, we used the default WRF version (v4.5.2), which supports only the Fitch scheme for WFP, and this can only be combined with the MYNN PBL scheme. The part below has been added to the revised text, in the conclusions, in lines 635-639:

The scope of this study is limited to using the Fitch WFP scheme combined with the MYNN 2.5 PBL scheme. Additional multi-year simulations using different WFP schemes, like the explicit wind farm (EWP) parametrization scheme [Volker et al., 2015, García-Santiago et al., 2024] applied with different PBL schemes or the induction correction to the original Fitch scheme [Vollmer et al., 2024], could also

provide insights into the year-long effects of large wind-farm clusters.

RC: 16. Line 6: *Spell out “1 km” as “one kilometer”.*

AR: We have corrected this in the revised text.

RC: 17. Line 118: *“...is the number of turbines...”*

AR: This has been changed in the revised version.

2. Reviewer #2

2.1. General comments

RC: *This paper evaluates the long-term capability of the WRF–Fitch WFP modeling framework representing wake effects from an offshore wind farm cluster in the Belgium–Dutch cluster. About two years of profiling lidar observations at several coordinates and SAR data from four selected dates are used for comparison.*

There is indeed a literature gap on the evaluation of the cluster wakes simulated by WRF-WFP over more realistic meteorological conditions using observations. This is no easy task, as it is difficult to cover large areas for a long period of time with measurement devices. Thus, the research problem and gap are well stated. The text is very well written, and the figures are well made; some of them convey information very nicely and look great (Figs. 14 and 15). The methodological approach makes sense to me, with some reservations. First evaluate the error metrics in detail to support the subsequent analyses of the wake effects.

However, my main concerns are the superficial interpretation of results, the limited engagement with the literature, and consequently, very mild conclusions for a study with such strong potential. The paper often remains at a descriptive level (“the wake extends”, “bias improves”) without explaining why certain behaviors occur or how they compare to previous WRF–WFP, LES, or observational studies. If the authors strengthen the physical interpretation of their results and relate them more thoroughly to existing literature, the manuscript could make a significant contribution.

AR: Dear Reviewer,

We sincerely thank you for your time and for the detailed, insightful comments. We appreciate the recommended references. We have included them in the revised version of the manuscript, as they align well with our current study. All changes and additions in the revised manuscript are highlighted in blue and enclosed in boxes.

We should also highlight that we’ve corrected the interpolation at hub-height for the BSB-lidar cases, resulting in a more accurate representation of the metrics throughout the full manuscript. More specifically, in the first version of the manuscript, we unintentionally used the lidar-measured heights without adding the BSB platform height of 45 m. This led to an incorrect estimation of both wind speed and wind direction at the intended hub height of 107 m.

2.2. Specific comments

RC: *1. Please provide details on the turbine types used, including hub height, rotor diameter, and ideally thrust/power coefficient curves. These curves can be placed in the Appendix. If some data are confidential, clarify which components are public and which are not.*

AR: A detailed table with types of turbines, hub-height, and rotor diameter has been added in the Appendix. Most of the power and thrust curves remain confidential.

RC: *2. A more detailed characterization of the sources of model error needs to be made. The paper clearly shows that the simulations with the Fitch WFP lead to improved metrics of wind speed and wind direction based on the lidar observations. In Section 3.2.1, wind speed metrics for several lidar locations and models are shown. However, these error metrics are aggregated for all wind directions, and so they represent simultaneously the effect of the WF simulation representing the undisturbed wind and wake effects. This is a problem because wake effects can be misrepresented, both because of errors in the representation*

of the free wind and specificities to the WFP. This information is crucial to establishing what makes the model more or less accurate (inflow error, wake modeling error, spatial gradients errors). As we know, errors in the undisturbed wind speed propagate into errors in the wake effects. This piece of evidence will strengthen the discussions when a few selected cases are presented in the SAR evaluation (Fig. 15)

AR: We have overall updated the revised manuscript considering this comment. More precisely:

- Regarding misrepresentation in errors in the free-stream, in Section 3.2.1 (lines 294-301):

The bias score (Figure 5A) shows that both ERA5 and the NWF simulation exhibit a consistent negative bias at all sites except BSB, a consequence of the inherent ERA5 underestimation that propagates into WRF. The additional wind speed reduction introduced by the Fitch WFP leads to a slightly more negative bias in the WF run, although WRF generally aligns better with the lidar observations than ERA5. **Part of the improved bias in the NWF runs likely results from an error-cancellation effect with the negatively biased ERA5 input, whereas in the WF simulations, these errors become slightly amplified due to wake-induced deficits, but remain lower than those in ERA5.** At BSB, both ERA5 and the NWF runs show a positive bias, while the WF simulation produces a smaller negative bias, corresponding to an improvement of about 82% over NWF and 73% over ERA5 in the near-wake under the predominant southwesterly flow.

- Two extra figures on wind speed and wind directions bias per wind direction have been added, i.e., Figures 7 and 8, to better understand WRF performance at different wind sectors. Additionally, these paragraphs have been added in Section 3.2.1 (lines 324-340):

In Figure 7, the wind speed bias for each wind direction at each location is depicted, allowing a clearer assessment of the bias sources. Colored lines represent NWF, WF, and ERA5 datasets, and the dashed lines indicate zero bias. At WHi, EPL, and LEG where wakes due to the existence of the Belgian-Dutch cluster are observed, the NWF and WF simulations generally show similar distributions, except along the principal wake directions, where NWF performs slightly better. More specifically, at WHi the dominant wake influence arrives from the northeast, and around 45° the NWF bias is closer to zero. A similar pattern appears at EPL, and LEG, where NWF bias is closer to zero around 225° and 255° , respectively. Especially at EPL, the bias is close to zero between 190° and 225° in the NWF simulations. In all three cases, both WRF configurations reduce the negative bias found in the ERA5 data, highlighting that the Fitch scheme introduces an additional bias downstream of wind farm regions. Conversely, in the intra-farm region at BSB, the Fitch scheme provides a more accurate wind speed distribution across all wind directions, underscoring its importance in strongly waked environments. This behavior of the Fitch scheme, showing prolonged wakes and enhanced wind speed deficits at locations farther downstream and more accurate wind speed distributions closer to the turbines, has also been reported in previous studies [García-Santiago et al., 2024, Ali et al., 2023].

Similar to the wind speed bias per wind direction, Figure 8 presents the wind direction bias. Although the mean bias in Figure 6A suggests a closer agreement between the lidars and ERA5, WRF provides a better representation of the wind directions when the wind originates from the north to south-southeast (0 to 180°). Conversely, for westerly wind directions between 180 and 360° ERA5 performs better. At BSB, the mean bias is nearly identical between the WF runs and ERA5, but ERA5 shows a slightly closer agreement with the lidar measurements across all wind directions.

- Furthermore, in Section 3.2.3, regarding error metrics under different wind speed regimes for SW and NE wind directions, Tables 4, and 5 have been added. These Tables present bias and cRMSE for the NWF and WF wind speeds against the three analyzed lidars under different wind directions and wind speed classes. Please see the response to comment 4, where the updates in the manuscript are discussed.
- In addition, in response to comment 11, mean absolute error (MAE) and absolute error (AE) for the SAR-events are presented in Section 3.3 (please have a look at comment 11).

RC: *3. Many more discussions are needed. For instance, Section 3.3 only describes the results without a single reference to the literature. That is unfortunate, because Figures 14 and 15 are the apex of the paper in my opinion. It is a missed opportunity to compare your evidence to the literature and describe the contrasts. Please cite papers that did WRF-WFP vs. lidar and/or SAR observations. Resort to the good references you already have and consider including Platis et al. (2018) and Cañadillas et al. (2020), which are great papers on the same scope. You cited Ali et al. (2023) in the Introduction, but their work is never mentioned again. The reader wants to know how the representation of the wake effects compares with other studies that use idealized LES or mesoscale simulations with observations near offshore wind farms.*

AR: We have updated the results sections throughout, incorporating additional literature and further details as requested. More precisely:

- Regarding Section 3.3, we have improved that part, linking the results and conclusions to existing literature. More precisely, we have included additional discussion on the stability stratification assessed in WRF simulations, and how it relates to the wake patterns, as well as with additional studies.

Figure 16 presents the comparison of SAR and WRF wind fields at 10-meter height. The left column shows the SAR-derived wind speeds, denoted by the subscript 1. The middle column depicts the WF wind speeds, indicated by a subscript 2, and in the right column, the reference NWF wind speeds are illustrated, and denoted with a subscript 3, as a reference case. In case A, higher wind speeds are also related to stable-neutral conditions and a shallow boundary layer (see Table 7). The wake structure of the Belgian-Dutch wind farm cluster is well captured by the WF simulation. However, a slight deviation in wind direction is observed in WRF, exhibiting a positive bias toward more westerly flow. Similar bias has also been highlighted in the wind roses (Figure 4) and in the wind direction validation heatmap in Figure 6A. The prolonged wake extending more than 50km downstream under stable-neutral stratification agrees with the findings of Cañadillas et al. [2020] and Platis et al. [2018], who also report that wake structures are more pronounced and longer during stable conditions. Furthermore, Siedersleben et al. [2020] examined a stable event simulated with the Fitch scheme in WRF (version 3.8.1), comparing the original formulation ($\alpha = 1$) with a variant excluding the added TKE term, and showed that including the added TKE source under stable stratification results in a more realistic representation of the prolonged wake of a SAR-retrieved 10-meter wind speed pattern. In our study, although we use a finer grid spacing (1km instead of 1.67km) and a reduced TKE coefficient ($\alpha = 0.25$), which limits the influence of the added TKE term, we observe a very good agreement between the SAR image and the WF simulations during stable conditions. Furthermore, Ali et al. [2023] evaluated several WFP schemes to characterize the wake of a specific event using 10 m SAR data. Their results show that, when comparing the Fitch model with $\alpha = 1$ and $\alpha = 0.25$, the latter produces a longer wake downstream of the wind farms, similar to the behavior observed in case A analyzed in the present study.

In case B, the wake is weaker, but still detectable both in SAR image and in the WF simulation. In this case, stability conditions are very unstable, and the PBLH is around 700 meters, explaining the attenuated wake pattern. Case C is characterized by higher wind speeds, L varies from -300 to -450 m, and PBLH is approximately 300 m across the three locations, resulting in a more unstable-neutral stratification profile, resulting in the observed downstream wake of about 40 to 50km. [Similar wake patterns but slightly shorter, about 25 to 35km have been observed during neutral stratification in the study of Platis et al. \[2018\].](#) Last, in case D, very unstable ABL conditions prevail, and the PBLH varies between 250 meters upstream at the WHi location to 358 meters downstream at the EPL area. This is the only SAR image with clear wakes for SW wind conditions. Overall, the patterns of the wakes are accurately represented by the WF simulations, highlighting the importance of using the Fitch WFP scheme in WRF. [Although SAR assumes neutral stability to compute the 10-meter wind speed, there are discrepancies with the WRF model, especially in cases B and D, where very unstable stratification dominates based on the values of Obukhov length. Uncertainty exists regarding the assumption of SAR image neutrality used to calculate wind fields, an issue that has also been reported in previous studies across different applications Hasager et al. \[2011a\], Ahsbahs et al. \[2017\]. Furthermore, there is uncertainty in the estimation of \$L\$ within WRF, as it strongly depends on the PBL and surface schemes applied in each configuration.](#)

In addition to the wind speed deficit maps, Figure 17 presents the transects that connect WHi, BSB, and EPL. The green line illustrates the 10-meter height SAR wind speeds, while the purple and orange lines correspond to the WF and NWF simulation outputs, respectively. Although a Hampel filter (see section 2.3.3) has been applied to the transect of the SAR data, we can still observe a few outlier values, especially in the intra-farm area. [The wind turbines generate spikes in the wind fields due to strong radar reflection \[Hasager et al., 2015, Ahsbahs et al., 2020\].](#) Therefore, to further evaluate the performance of the WRF model, hub height-lidar measurements are incorporated when available, depicted as gray scatter points in Figure 17.

- Section 3.2.1 has also been updated, providing more details in the error-sources and links to literature (e.g., García-Santiago et al. [2024], Ali et al. [2023]), as described in the reply to comment 2.
- Section 3.2.3 has also been improved according to the next comment 4. Specifically, Tables 4, and 5 have been added, where wind speed bias and cRMSE metrics for the NWF and WF simulations are discussed (linked to comment 2 as well), and the discussion on Figures 12 and 13 has been updated and linked to existing literature (e.g., Ali et al. [2023], Fischereit et al. [2024a]). In the same Section, Figure 14 has been improved, including also a classification of averaged wind speed deficits under different wind speed magnitudes and stability conditions, for the analyzed SW, and NE wind directions, showing that an increase in the wake magnitude is observed when transitioning from unstable to neutral and stable conditions (see comment 7).

RC: *4. Following on the aforementioned point, Figures 10 and 11 are not being discussed as well, which is another missed opportunity to communicate whether your results make sense with the literature or not.*

AR: The updated Figures 12, and 13 (i.e., Figures 10, and 11 in the first version of the manuscript) are now thoroughly analyzed and discussed. Fu, Tables 4, and 5 have been added, where wind speed bias and cRMSE metrics for the NWF and WF simulations are discussed. Therefore, we improved the clarity and the flow of this Section. Please find below the updates and changes in Section 3.2.3 in lines 398-433:

Figures 12 and 13 depict the wind speeds along the transect connecting WHi, BSB, and EPL for the four different wind speed classes under SW and NE wind directions, respectively. The distributions at each location are clipped at the 1st and 99th percentiles solely for illustrative purposes, without affecting the presented statistics. Overall, the WF simulations slightly underestimate the wind speed, and this is corroborated by the bias scores in the full-period heatmaps presented in Figure 5A. More precisely, under SW winds, the model underestimates wind speeds at the upstream location of WHi across all the wind speed classes. Especially, when the wind speed is between $[3, 5) \text{ m s}^{-1}$ or above 8 m s^{-1} , the produced wind speed deficit of the Fitch model is overestimated. On the other hand, the model provides a more accurate wake pattern when wind speed belongs to $[5, 8) \text{ m s}^{-1}$ and $[8, 11) \text{ m s}^{-1}$ classes. These wind-speed ranges usually correspond to the power ramp-up region of a wind turbine, where power output exhibits strong sensitivity to wind-speed variations. Under northeasterly winds (see Figure 13), the model underestimates wind speeds in both the $[3, 5)$ ranges across all locations, while performance improves markedly when wind speed varies from 5 to 11 m s^{-1} . In both wind direction cases, the WF simulations overestimate the intra-farm wind speed deficit when wind speed is above 11 m s^{-1} .

Table 4 summarizes the wind speed bias and cRMSE across the wind speed classes for the SW wind direction, highlighting how model performance varies in upstream, intra-farm, and downstream locations. Under SW winds, where the flow progresses from WHi (upstream) $\rightarrow$ BSB (intra-farm) $\rightarrow$ EPL (downstream), the largest errors occur upstream at WHi in the $3\text{--}5 \text{ m s}^{-1}$ class (bias= -1.09 m s^{-1} , and cRMSE= 1.56 m s^{-1} , which are very similar to WF and NWF runs), reflecting that small inflow uncertainties propagate strongly before the flow encounters the farm. As the flow moves into and through the farm, the bias systematically decreases, reaching nearly zero at BSB and EPL in the $5\text{--}8 \text{ m s}^{-1}$ class in the WF simulations. The cRMSE increases with the wind speed magnitude, indicating that although the mean wind speed is well captured, the variability associated with wake turbulence and flow transitions is still not well reproduced. Overall, the cRMSE in the WF simulations is comparable to, or slightly lower than, that in the NWF simulations across all locations and wind speed regimes.

Under NE winds, where the flow reverses (EPL $\rightarrow$ BSB $\rightarrow$ WHi), a similar pattern emerges as presented in Table 5. The upstream location (now EPL) shows a reduced bias at low wind speeds (about -0.50 m s^{-1} in both NWF and WF) but a higher cRMSE (1.8 to 1.95 m s^{-1}). Within the farm (BSB), bias remains modest across most wind speed classes, while cRMSE stays elevated. When wind speed is in the $8\text{--}11 \text{ m s}^{-1}$ class, WF provides an almost negligible bias both at EPL and BSB. At higher wind speeds (above 11 m s^{-1}), both bias and cRMSE increase, especially at BSB. The improved bias in the NWF simulations and under specific wind speed magnitudes can be associated with error-cancellation of the already biased ERA5 data (see section 3.2.1). The differences in bias and cRMSE scores between the NWF and WF simulations at the upstream locations (WHi under SW wind directions, and EPL when wind comes from NE) are related to the fact that the flow can already be slightly affected by the UK wind farms on the west and the Dutch cluster on the NE, as presented in Figure 1b.

Although Figures 12 and 13 present transects under different wind speed magnitudes for the SW and NE wind directions, transect-based analyses of instantaneous wake events have also been reported using flight measurements [Ali et al., 2023, Fischereit et al., 2024b]. Those studies showed that the Fitch WFP in WRF can reproduce the wake features and reduce wind speed biases along transects. The transects in Figures 13 and 14 extend this type of analysis but rely on lidar observations rather than flight data, allowing for a comparison of more events and analyzing upstream, intra-farm, and downstream wake characteristics.

RC: 5. *Figure 11: You state that “Notably, intra-cluster wake behavior differs under NE winds, with a continuous reduction in wind speed observed throughout the wind farm”, and no explanation as to why this behavior takes place is provided. Why is it that the wind speed deficit is more flattened for the SW direction? As readers, we want these insights from you. Sometimes the most interesting aspects of the research appear in these details. What is the role of the different transects relative to the wake center? What is the role of the heterogeneous layout and density of the wind turbines?*

AR: There are two main reasons that we observe different wake structures on the SW and NE wind directions, and they are both related to the geometry of the Belgian Dutch cluster. First, in the SW case, the flow meets the Belgian cluster that extends perpendicular to the flow, and it has a dense wind turbine-layout. This results in a more rapid reduction of the wind speed, as well as the development of a blockage zone upstream of the Belgian cluster. The already reduced wind speed remains almost constant downstream of the Belgian cluster and when crossing the Borssele wind farm in the Netherlands. On the other hand, when the wind direction is NE, the reduction of the wind speed is smoother, as the Borssele wind farms cluster is less dense and has a greater area. When crossing the Belgian zone, an additional wind speed reduction is observed. This paragraph has been improved, including also the maximum wind speed deficits in the updated figures 12 and 13, and this statement has been included in the revised version of the manuscript in lines 434-452:

Notably, distinct intra-cluster wake patterns emerge under SW and NE wind conditions. Under SW winds, the strongest wind speed deficit originates within the Belgian wind farm cluster and remains nearly unchanged as the flow proceeds across the Borssele zone. In contrast, under NE winds, the wind speed decreases progressively throughout the entire wind farm region, beginning in the Borssele cluster and intensifying as the flow enters the denser Belgian wind farms. Furthermore, the wake recovery rate is less smooth at high wind speed magnitudes. These findings highlight the directional dependence of wake dynamics and underscore the need for improved model representation of low-speed flows and intra-cluster wake effects. More specifically, the maximum wind speed deficits increase with wind speed magnitude. At wind speeds above 11 m s^{-1} , the difference between SW and NE conditions becomes substantial: under SW winds, the maximum deficit reaches approximately 2.72 m s^{-1} , whereas under NE winds it increases to about 4.19 m s^{-1} . At lower wind speeds (below 8 m s^{-1}), the largest deficits in both SW and NE regimes occur near the center of the densely built Belgian wind farm zone. At higher wind speeds, however, the location of the maximum deficit shifts. Under SW winds, the strongest deficit moves toward the northeastern boundary of the Belgian zone, near the interface with the Dutch Borssele cluster. Conversely, under NE winds, the maximum deficit appears at the downstream exit of the Belgian wind farm cluster. These contrasting behaviors between the wake trends under SW and NE wind directions arise from the strong dependence of wake evolution on turbine spacing. The averaged density of the Belgian wind farms that the analyzed transect crosses is about 10 to 12 MW/km^2 , while the Borssele cluster has a density of about 4 to 5 MW/km^2 . Therefore, the dense Belgian cluster produces rapid wake merging and a deeper cumulative deficit, whereas the more widely spaced Borssele layout allows partial recovery before wakes redevelop downstream. As a result, SW winds encounter an already merged wake that weakens over the coarse Borssele area, while NE winds first traverse the coarse Borssele layout, where deficits remain modest, and then intensify sharply upon entering the dense Belgian cluster.

RC: 6. *L467–470: “Systematic biases in both wind speed and direction are evident, particularly in the wind speed metrics. However, it should be highlighted that the model is driven by ERA5 hourly data, which consistently introduce a negative bias, affecting the boundary conditions in WRF and propagating wind speed underestimations to the analyzed locations.” And what is the effect of this modeling bias on the representation of the wake effects? This discussion is fundamental because sometimes you have a bias in*

the free wind speed that can actually improve, by accident, the validation metrics. This must be explained.

AR: This has been partially addressed in comment 2. The negative bias in ERA5 can affect the wake assessment. It is shown in Section 3.2.1, in Figures 5 and 7, where NWF can provide lower biases compared to the WF runs (not at highly-waked locations like BSB). We consider that this can be due to an error-cancellation effect, as the ERA5 is already negatively biased.

This paragraph in Section 3.2.1 (lines 294-301) has been added:

The bias score (Figure 5A) shows that both ERA5 and the NWF simulation exhibit a consistent negative bias at all sites except BSB, a consequence of the inherent ERA5 underestimation that propagates into WRF. The additional wind speed reduction introduced by the Fitch WFP leads to a slightly more negative bias in the WF run, although WRF generally aligns better with the lidar observations than ERA5. **Part of the improved bias in the NWF runs likely results from an error-cancellation effect with the negatively biased ERA5 input, whereas in the WF simulations, these errors become slightly amplified due to wake-induced deficits, but remain lower than those in ERA5.** At BSB, both ERA5 and the NWF runs show a positive bias, while the WF simulation produces a smaller negative bias, corresponding to an improvement of about 82% over NWF and 73% over ERA5 in the near-wake under the predominant southwesterly flow.

In addition, in the analysis of SW, and NE wind direction in Section 3.2.3, in Figure 12, and Table 4, we observe that in the upstream location (WHi under SW wind conditions), there is always a negative bias, most likely propagated from ERA5 and the overall WRF performance. This under specific conditions, for example, in Table 4, under [3-5) m s⁻¹ wind speeds, BSB provides a negligible bias in the NWF simulations that is higher in the WF. We consider that this can happen due to error cancellation. Similarly, under NE wind conditions in Table 5, when the wind speed is low, the bias at BSB is lower in NWF simulations compared to the WF.

In the conclusions, this paragraph (lines 593-605) has been updated as well:

Our findings demonstrate that the use of the Fitch WFP scheme, in combination with this particular WRF configuration, captures the wake effects and provides an improvement in model performance relative to a simulation without wind farms at the intra-farm region. First, focusing on the time-averaged wind speeds over the full analysis period, the WF simulations yield a substantial improvement in both bias and EMD, with reductions of approximately 82% and 73% compared with the NWF simulations, respectively, in the intra-cluster (BSB location). **Furthermore, when analyzing the wind speed bias under different wind directions, both NWF and WF configurations reduce the negative bias inherited from the ERA5 data, with the results indicating that the Fitch scheme introduces an additional deficit downstream of the Belgian-Dutch cluster.** However, within the intra-farm area, the Fitch scheme provides a more accurate wind speed distribution across all wind directions, underscoring its importance for representing strongly-waked environments. Beyond the improvements in bias and EMD, the WF simulations perform similarly to the NWF runs in terms of cRMSE and correlation, with only small and marginal differences. Similarly, wind direction metrics are also improved in the WF simulations, as bias is reduced by approximately 51% at BSB, while an analysis of wind direction biases by sector reveals that WF can perform better than ERA5 for wind directions from northeast to south.

RC: *7. Many figures address modeling error metrics (Figs. 5–8), but none for the actual wake effects. I think we can increase the value of the paper if you use Figure 13 as a template and plot a wake-related metric*

(e.g. wind speed deficit). That template is great because it simultaneously separates the effects of wind speed classes and stability.

AR: The manuscript has been updated to include wake-effects metrics in terms of wind speed deficit magnitude, as well as wake area and length. More precisely, Figure 14 has been moved to Section 3.2.3, and it has been updated, including also the average wind speed deficit magnitude per wind speed magnitude class for each ABL stability regime. Below in bold (lines 462-477) is presented the paragraph describing wind speed magnitudes under different stability regimes and wind speed conditions.

Figures 14A, and B present the frequency of the ABL stability conditions under the different wind speed ranges, for SW and NE wind directions, respectively. The ABL stability classification is based on Obukhov length, averaged across the three lidar probes in the NWF simulations. Under low wind speed conditions, very unstable and unstable events tend to dominate. Conversely, a transition toward near-neutral ABL stratification under high wind speeds is also evident. Previous studies [Rosencrans et al., 2024, Palatos-Plexidas et al., 2024, Porchetta et al., 2024] have demonstrated the dependence of wake propagation and wind speed losses on ABL stability, showing that transitioning from neutral to stable and very stable stratification, wakes tend to extend further downstream of wind farms. Similarly, we underscore the relation between wind speed magnitude, ABL stratification, and wake propagation, where during high wind speeds, the frequency of neutral events is enhanced and wakes propagate a few tens of kilometers downstream of the Belgian-Dutch cluster.

In addition to the frequency of the stability regimes, in Figures 14C and D, the magnitude of the average normalized wind speed deficit is presented under the different stability conditions. The wind speed deficits are estimated as the average across the three analyzed lidar locations. It is observed that under very unstable and unstable occurrences, the wake magnitude is relatively lower (varying from 4 to 8%). When ABL stratification is near neutral, higher wake magnitudes are also observed. Following the patterns presented in studies [Rosencrans et al., 2024, Palatos-Plexidas et al., 2024, Porchetta et al., 2024], moving from unstable to neutral and stable stratification results in higher wind speed deficits. However, this should be interpreted together with the frequency of events in each wind speed class, which shows that the number of very stable cases, where wake effects are relatively strong, is not particularly large. On the other hand, the wind speed deficits during very unstable and unstable events that occur more frequently (especially when wind speeds are below 8 m s^{-1}) are significantly lower. Near neutral conditions occur more frequently when wind speed is above 8 m s^{-1} in both SW and NE wind direction, and the wake magnitude in those cases ranges from about 7 to 12%. Greater wind speed magnitudes are associated with more pronounced downstream wind speed deficits (see Figures 12 and 13), as higher wind speeds in this region occur more frequently under neutral to stable conditions. Under these regimes, particularly during stable conditions, reduced turbulent mixing limits wake recovery, leading to enhanced and more persistent wakes [Rosencrans et al., 2024, Ali et al., 2023, Siedersleben et al., 2020]. The transition toward neutral and stable ABL conditions with the increasing wind speed may contribute to the higher cRMSE values observed in Tables 4 and 5, as these regimes are associated with enhanced shear.

Furthermore, in Figure 15, Section 3.2.4, the 95% wake area is now depicted with purple solid lines. The maximum wake length of the wake recovery area per direction is shown in purple dashed lines. The wake length and wake area metrics are also presented in Table 6. Lines 481-501 have been updated as follows:

The analysis presented in the previous section, based on SW and NE wind directions, supports the investigation of the two-dimensional wake characteristics and blockage effects within the Belgian-Dutch wind farm cluster. A reduction in wind speed upstream of the wind farms is evident both under SW and NE wind directions. In the transect under SW winds (Figure 12), a noticeable wind speed deficit begins approximately 8 km upstream of the wind farm. This reduction is present across the different wind speed magnitudes, but becomes more rapid around 4 to 5 km upstream, particularly when wind speeds exceed 5 m s^{-1} . For NE wind directions (Figure 13), the blockage effect is also visible, though it is confined to a smaller region, approximately 2 to 3 km upstream of the Borssele wind farm.

The wake maps in Figure 15 further confirm the presence of blockage effects under both SW and NE wind conditions. The solid purple lines indicate the 95% wake recovery area, while the dashed purple lines represent the maximum wake length within this area for each wind speed regime and wind direction. The wake structures become less coherent at wind speeds below 5 m s^{-1} while an induction zone defined by the 95% wake recovery area is observed in low wind speeds. More precisely, the maximum blockage distance from the front row of the wind turbines appears around 10 km under SW wind direction (Figure 15A₁), whereas around 5 km in the NE wind direction regime (Figure 15A₂). In addition, Table 6 presents the values of the 95% wake recovery area and the maximum wake length within this area. The largest coherent wake region is extracted using connected-component labeling. Its perimeter is obtained from the contour of the binary mask, and the maximum extent is computed as the convex-hull diameter of the perimeter points. The estimation of both the 95% wake recovery area and the maximum wake length near SW or NE wind directions is performed using the SciPy Python package[Virtanen et al., 2020]. The values under the different wind speed regimes in Table 6 are comparable for both SW and NE wind directions.

When the flow transitions to higher wind speeds (above 5 m s^{-1}), the wake structures become more coherent, with more pronounced wake recovery areas, leading to overall longer wake lengths (see Table 6). In the $5\text{--}8 \text{ m s}^{-1}$ class, a blockage region forms upstream of the turbines, extending approximately 8 to 9 km from the first row under both SW and NE conditions. At higher wind speeds, this blockage pattern becomes negligible, and the flow tends to deflect around the cluster and propagate farther downstream.

RC: 8. *On the marine ABL being mostly unstable: “Similar conditions have been mentioned in the work of Rosencrans et al. (2024) analyzing the ABL stratification on the Rhode Island–Massachusetts area in a WRF execution from September 2019 to September 2020.” Actually, Rosencrans et al. show that unstable vs. stable are almost 50-50, with a few occurrences of neutral cases. I think this statement needs an adjustment. Pointing out that your case is more unstable is actually interesting and adds to the discussion.*

AR: We have adjusted the text, as, although in the works of Archer et al. [2016], Porchetta et al. [2024], Palatos-Plexidas et al. [2024] there is a predominance of unstable events, in the study of Rosencrans et al. [2024], there is almost an equal number of unstable and stable events with a few neutral occurrences. Therefore, this paragraph has been changed in the manuscript as follows in lines 268-274

Overall, the predominance of unstable to very unstable events has been highlighted in several studies over the past decade. For example, Archer et al. [2016] demonstrated that the marine ABL along the U.S. northeastern coast was mostly unstable during a 2003-2011 measurement campaign. Similar conclusions for the North Sea have been reported in recent WRF-based studies [Porchetta et al., 2024, Palatos-Plexidas et al., 2024]. However, it is important to note that ABL stability strongly depends on

the region and period of interest. For instance, Rosencrans et al. [2024] found an almost equal mix of unstable and stable conditions, with relatively few neutral cases, in the Rhode Island-Massachusetts area based on WRF simulations from September 2019 to September 2020.

RC: *9. You mention the Vollmer et al. correction in the introduction, but it was not used in the paper. Can you comment on what effect the Vollmer et al. correction would have on your simulations?*

AR: In the paper of Vollmer et al. [2024], the authors conclude that using the induction-correction method for multiple turbines within a computational cell (i.e., 5 22MW Turbines in that case, in the Fitch-mAIF model) can reproduce much better the power curve combined with the original Fitch implementation. On the other hand, they also highlight that downstream wind speeds applying the original Fitch scheme in WRF simulations have shown good agreement with measurements [Cañadillas et al., 2022, Fischereit et al., 2022], but this may occur partly because the Fitch scheme ignores induction effects and inner-grid wake interactions when multiple turbines occupy the same grid cell. Instead of neglecting these processes, the reduction in thrust and power from turbine clusters due to sub-grid wake effects should be estimated using an induction-corrected reference wind speed. The proposed correction is therefore only a first step toward a more physically consistent wind farm parameterization in mesoscale models. We should also highlight that in the work of Vollmer et al. [2024] a 2km grid resolution is used on a 5-day-long simulation.

Although we should take into account that the correction in the induction zone in the Fitch model proposed by Vollmer et al. [2024] has not been thoroughly studied yet, and further validation analysis is needed across longer periods, the proposed method is very promising and should be considered for future year-long WFP studies.

Considering that this three-year-long 1km study focuses only on the original Fitch WFP, coupled with the MYNN PBL scheme, we have included this in the conclusion for future studies: The part below has been added to the revised text, in the conclusions, in lines 635-639:

The scope of this study is limited to using the Fitch WFP scheme combined with the MYNN 2.5 PBL scheme. Additional multi-year simulations using different WFP schemes, like the explicit wind farm (EWP) parametrization scheme [Volker et al., 2015, García-Santiago et al., 2024] applied with different PBL schemes or the induction correction to the original Fitch scheme [Vollmer et al., 2024], could also provide insights into the year-long effects of large wind-farm clusters.

RC: *10. The SAR analysis looks great. But why have only 4 dates been selected over 2 years?*

AR: SAR images cannot easily capture wind farm wakes, as they primarily provide wind speeds at approximately 10m height. In addition, only about 8–10 SAR snapshots per month cover the area of interest, and only a few of these contain sufficiently clear wake signatures. Furthermore, we were specifically interested in SW and NE wind directions affecting the Belgian–Dutch cluster, as these can generate wakes that propagate toward the WHi (located SW the Belgian–Dutch cluster) and EPL (positioned NE the Belgian–Dutch cluster) lidar sites. As a result, the number of suitable SAR snapshots becomes even more limited. For this reason, we carefully selected the snapshots with clearly detectable long wake patterns (rather than short downstream deficits) to evaluate the WRF-WF performance during those cases. The snapshots are illustrative and may therefore reflect selection bias.

The overall text in Section 2.3.2, has been updated as follows (lines 191-201):

The SAR data used in this study were retrieved by the Sentinel-1 European Space Agency mission. More specifically, there are two satellites, Sentinel-1A (2014-present) and Sentinel-1B (2016-2021), equipped with C-band Synthetic Aperture Radar (SAR) instruments operating at 5.405 GHz, and both satellites follow a sun-synchronous polar orbit at an altitude of 693 km [Hasager et al., 2024]. Focusing on the area of interest, including the Belgian-Dutch wind farms cluster in the Southern Bight of the North Sea, as well as additional upstream and downstream areas to study potential wake structures, the analyzed time instances occur approximately at 06:00 UTC every twelve days, and at 17:40 UTC every six days, respectively (Sentinel-1A, 2021-2023). The operation of Sentinel-1B during 2021 provided an additional set of overpasses at around 17:32 UTC every six days. This corresponds to approximately 8 to 10 SAR snapshots per month over the area of interest during the analyzed period. The wind speed at 10-meter height is then estimated by applying the CMOD5N geophysical model function [Hersbach et al., 2007], as explained also in the work of Siedersleben et al. [2020]. This indirect retrieval method assumes that radar backscatter from the sea surface, which reflects centimeter-scale waves generated by instantaneous wind stress, can be converted into wind speed at 10-meter height [Ahsbahr et al., 2017, 2018]. The wind fields are provided at a 500-meter pixel resolution through the Global Wind Atlas Science Portal of the Danish Technical University (DTU) Wind Energy Department (<https://science.globalwindatlas.info/#/map/satwinds> last access: 30 July 2025).

This has also been updated in Section 3.3., lines 504-510

Four specific timestamps have been selected to analyze the wind speed patterns derived from SAR and WRF data, as well as to validate the model performance across the transect that connects the WHi, BSB, and EPL lidars (see Figure 2). These timestamps were selected manually based on the visual appearance of wind farm wakes in the SAR images. Although an average of 8 to 10 snapshots per month (see section 2.3.2) covers the analyzed region, the focus on wake events affecting the WHi, BSB, and EPL lidar sites under SW and NE wind directions resulted in a further reduction of the available snapshots. Therefore, an additional event from September 2020, outside the defined three-year simulation period, has been deliberately included for analysis.

RC: *11. Are the biases of SAR observations fixed or vary in space and time? Can you estimate the SAR bias at 10 m for your case? The agreement does look very good in the wake region (errors of about 0.5 m/s or less at times), but more information on the reliability of these observations, preferably citing the literature, can strengthen your claims that the model works well.*

AR: We have assessed the mean absolute error between WRF and SAR at 10 meters height across the transects presented in the updated Figure 17 in the revised text. In addition, absolute error (AE) is provided between WRF and lidar measurements (when available). This has been added in the revised text in Section 2.3.4 (lines 242-245):

Last, for the SAR transects, we compute the mean absolute error, $MAE = \sum_1^n |U_{SAR} - U_{WRF}|$, where n is the number of points across the upstream, intra-farm, and downstream regions of each transect (70 points in total after interpolation of the SAR and WRF data). In addition, for the same time instances, we evaluate the point-wise absolute error, $AE = |U_{lidar} - U_{WRF}|$, using the lidar observations as reference.

The results of the evaluation for the SAR time instances are presented in Section 3.3., in Table 8. Furthermore, the overall text in Section 3.3., has been improved to include the metrics presented in Table 8, as well as a

more detailed discussion on the SAR-detected wakes according to comment 3. More precisely, regarding the validation of WRF against the four SAR-retrieved occurrences, the first paragraph of Section 3.3., has been updated as follows (lines 512-517):

In this section, we focus on a one-by-one comparison of the WRF model output with SAR data. SAR images are retrieved from Sentinel 1A and 1B satellites as explained in section 2.3.2, and they provide wind speed fields at 10-meter height. Four specific timestamps have been selected to analyze the wind speed patterns derived from SAR and WRF data, as well as to validate the model performance across the transect that connects the WHi, BSB, and EPL lidars (see Figure 2). These timestamps were selected manually based on the visual appearance of wind farm wakes in the SAR images. Although an average of 8 to 10 snapshots per month (see section 2.3.2) covers the analyzed region, the focus on wake events affecting the WHi, BSB, and EPL lidar sites under SW and NE wind directions resulted in a further reduction of the available snapshots. Therefore, an additional event from September 2020, outside the defined three-year simulation period, has been deliberately included for analysis. The exact datetimes and the ABL stratification conditions derived from the NWF simulations, based on the Obukhov length and the planetary boundary layer height (PBLH), are provided in Table 7. *Although this study uses only four specific SAR scenes and does not explicitly evaluate SAR data reliability, previous validation works provide strong evidence supporting the accuracy of SAR-retrieved winds. For example, Hasager et al. [2011b] compared SAR-derived wind speeds with meteorological mast measurements in the Baltic Sea and reported a small bias of -0.25 m s^{-1} and a correlation coefficient of $R^2 = 0.783$ across 900 images. Similarly, de Montera et al. [2020] assessed 1544 SAR Level-2 Ocean product instances against four offshore buoys and three coastal weather stations around Ireland, finding an average bias of -0.4 m s^{-1} .*

In the same Section 3.3., Table 8 has been added, and the text has been revised accordingly in lines 554-585:

Therefore, to further evaluate the performance of the WRF model, hub height-lidar measurements are incorporated when available, depicted as gray scatter points in Figure 17. In addition to the qualitative assessment of the wind speeds in Figure 17, Table 8 presents the MAE across the upstream, intra-farm, and downstream transect regions, as well as the point-wise AE, when lidar measurements are available, and the evaluation methodology is described in section 2.3.4. In cases A, B, and C (NE wind flow), the transect follows the direction EPL (upstream) → BSB (intra-farm) → WHi (downstream), whereas case D corresponds to a SW flow event, where the sequence reverses.

Case A shows good agreement in the representation of the transect wake. Specifically, while the WF simulation tends to underestimate wind speeds upstream and within the Borssele zone, it accurately captures the wind speed deficit in the Belgian wind farms and downstream of the wind farm cluster. This is also reflected in Table 8, where only in the upstream region WF generates a slightly higher MAE = 0.49 m s^{-1} , compared to MAE = 0.37 m s^{-1} in the NWF simulation. However, the improvement between the WF against the NWF simulations in MAE in the intra-farm region and downstream is 81% and 85%, respectively. Similarly, in Case B, both the intra-farm and downstream wind speed patterns are well represented by the WF simulation. Although the overall trend is accurately reproduced, the WF simulation tends to overestimate the wind speeds upstream and downstream of the wind farms. Additionally, a flow acceleration is observed as the wind exits the Belgian-Dutch cluster. This pattern is consistent with the broader flow conditions downstream of the Belgian-Dutch cluster, since a similar acceleration is evident in the SAR observations and in both the WF and NWF simulations. Both the lidar measurements and the WF simulation effectively capture the intra-farm wind speed deficit and the

downstream flow acceleration. Furthermore, in Table 8, MAE and AE are better in the WF than NWF simulation across the full transect regions.

Case C also features NE incoming wind conditions. Similar to the previous cases, the WF simulation successfully captures the wind speed deficit trends both within the intra-cluster zone and in the downstream wake recovery region. Both the NWF and WF simulations show noticeable discrepancies upstream at EPL compared to lidar observations with an AE reaching above 3 m s^{-1} . This is significantly improved both at BSB and downstream at WHi, where WF simulations outperform the NWF runs both in terms of MAE across the transect regions and AE against the lidar measurements. During this event, the Obukhov length at EPL is relatively high and negative, indicating unstable to near-neutral conditions. This may be attributed to less negligible shear, which can contribute to increased deviations from the lidar measurements. In the NWF simulation, an unrealistic and nearly linearly increasing pattern in wind speed of approximately 0.5 m s^{-1} between EPL and WHi is observed. Case D is the only selected timestamp with a SW wind direction. Although the SAR data exhibit considerable noise, particularly within the intra-farm region, the WF simulation is still able to reproduce the wind speed deficit and the downstream wake-recovery trend. In this case, only EPL lidar measurements are available, which show good agreement with the WF simulation, with a slightly improved AE compared to the NWF run. Furthermore, the MAE provides a more accurate matching with the 10-meter SAR wind speeds across the three transect regions in the WF simulations.

RC: *12. In L367–368 you state that “greater wind speed magnitudes are associated with more pronounced downstream wind speed deficits”. Why does the wind speed deficit increase with the wind speed in Figures 10–12?*

AR: The increase in wind speed deficit with increasing wind speed magnitude is primarily linked to the associated atmospheric stability conditions. As shown in Figures 14A, 14B, higher wind speeds (above 8 m s^{-1}) are more frequently associated with near-neutral stratification in both SW and NE flow regimes. Figures 14C and 14D further indicate that neutral and stable conditions correspond to larger normalized wind speed deficits. This behavior is consistent with previous studies, which show that neutral and stable atmospheric stratification promotes enhanced and more persistent wakes (e.g., Rosencrans et al. [2024], Palatos-Plexidas et al. [2024]). We have clarified this point in the revised manuscript in lines 472–477.

Greater wind speed magnitudes are associated with more pronounced downstream wind speed deficits (see Figures 12 and 13), as higher wind speeds in this region occur more frequently under neutral to stable conditions. Under these regimes, particularly during stable conditions, reduced turbulent mixing limits wake recovery, leading to enhanced and more persistent wakes [Rosencrans et al., 2024, Ali et al., 2023, Siedersleben et al., 2020]. The transition toward neutral and stable ABL conditions with the increasing wind speed may contribute to the higher cRMSE values observed in Tables 4 and 5, as these regimes are associated with enhanced shear.

RC: *13. The Conclusions need an upgrade after the description and discussion of the results. As of now, it states that “Our findings demonstrate that the use of the Fitch WFP, in combination with this particular WRF configuration, is largely able to capture wake effects and improve the general accuracy over a simulation without wind farms included.” Then, you proceed to describe that “The WF simulations accurately capture the wake structure when wind speeds range between 5 to 11 m/s.” As you can see, in the Conclusions, you merely restate the results. In my view, this restatement of results occurs because of the insufficient description and discussion of the results, as I said before.*

AR: The analysis and the discussion in the results Sections, and the conclusions have been updated overall. For

instance, lines 593-605 have been changed accordingly.

Our findings demonstrate that the use of the Fitch WFP scheme, in combination with this particular WRF configuration, captures the wake effects and provides an improvement in model performance relative to a simulation without wind farms at the intra-farm region. First, focusing on the time-averaged wind speeds over the full analysis period, the WF simulations yield a substantial improvement in both bias and EMD, with reductions of approximately 82% and 73% compared with the NWF simulations, respectively, in the intra-cluster (BSB location). Furthermore, when analyzing the wind speed bias under different wind directions, both NWF and WF configurations reduce the negative bias inherited from the ERA5 data, with the results indicating that the Fitch scheme introduces an additional deficit downstream of the Belgian-Dutch cluster. However, within the intra-farm area, the Fitch scheme provides a more accurate wind speed distribution across all wind directions, underscoring its importance for representing strongly-waked environments. Beyond the improvements in bias and EMD, the WF simulations perform similarly to the NWF runs in terms of cRMSE and correlation, with only small and marginal differences. Similarly, wind direction metrics are also improved in the WF simulations, as bias is reduced by approximately 51% at BSB, while an analysis of wind direction biases by sector reveals that WF can perform better than ERA5 for wind directions from northeast to south.

In lines 611 to 617, there is an updated version in the conclusions:

Extending the performance metrics to wake characteristics under different wind speed magnitudes, and focusing on SW and NE wind directions, the analysis revealed several consistent patterns. First, under SW wind direction, the WF simulations accurately capture the wake structure when wind speeds range between 5 to 8 m s⁻¹, providing an almost negligible bias at BSB and downstream at EPL. Under NE wind direction, the optimum performance of WF simulations occurs at wind speeds between 8 and 11 m s⁻¹. Although the model tends to overestimate intra-farm wind speed deficits at BSB under higher wind speed conditions, it shows comparable improvements in long-distance downwind wakes at EPL (for SW winds) and WHi (for NE winds), where smaller bias and cRMSE values are obtained.

And in lines 623 to 634:

Furthermore, a comparison with SAR images at specific events shows that WF performs well in representing the wake structure generated by the Belgian-Dutch cluster. Both in the intra-farm and the wake recovery regions, WF provides a realistic wind speed deficit. To assess the performance of the model during these events, MAE and AE are computed. More specifically, MAE is estimated across the upstream, intra-farm, and downstream regions, while AE is calculated at each lidar location when measurements are available. Building on the previously discussed improvements in the qualitative wake-pattern representation, the error metrics further confirm that the WF simulations provide a better alignment with observations than the NWF runs across all locations, with the most pronounced improvements occurring in the intra-farm and downstream regions.

This study can accurately evaluate the model performance by utilizing upwind, intra-farm, and downwind lidar measurements as well as SAR images at specific events. Systematic biases in both wind speed and direction are evident, particularly in the wind speed metrics. However, it should be highlighted that the model is driven by ERA5 hourly data, which consistently introduce biases, affecting the boundary conditions in WRF and propagating wind speed underestimations to the analyzed locations.

RC: *14. In Table 2, you use a stability classification that includes classes such as “stable–neutral”. Whereas this is valid, most commonly you will find a five-class system with “neutral”, “weakly stable”, and “very stable”. Perhaps using the five-class system will make it easier to discuss your work with the literature and for other researchers to interpret your results. This is not criticism, just something to be aware of. Your work, your choice.*

AR: Thank you, in our study, we use an adapted classification based on the study of Sathe et al. [2011]. We will consider using the suggested ABL stability classification in future works.

Technical corrections

RC: *1. Figure 1 needs domain dimensions in degrees or km.*

AR: We have updated Figure 1 in the revised text.

RC: *2. The L parameter is just Obukhov length, not Monin-Obukhov length, even though it belongs to the Monin-Obukhov Similarity Theory. This requires changes in Tables and in the text e.g. when you write “MO length”.*

AR: We have addressed this in the revised text.

RC: *3. Please include a wind direction arrow in Fig. 15, as you did for Fig. 11.*

AR: Figure 17 (in the revised text) has been improved accordingly.

3. Reviewer #3

RC: *The paper presents a three-year WRF mesoscale simulation study of a Belgian–Dutch offshore wind-farm cluster, with and without the Fitch wind-farm parameterization. The simulations are validated using three offshore lidars, and model performance is evaluated for different stability conditions. The main question the paper addresses is the impact of the wind-farm parameterization on the results. The authors find that including the wind-farm parameterization in the mesoscale simulations improves accuracy in both intra-farm and downstream wakes.*

Overall, this is a well-designed study that uses quantitative metrics to assess performance across different boundary-layer stabilities. The addition of SAR data alongside the lidars is a clear strength. I have a few general comments and questions I would like the authors to address before I can recommend this work for publication.

AR: Dear Reviewer,

We sincerely thank you for your time and for the detailed and insightful comments. All changes and additions in the revised manuscript are highlighted in blue and enclosed in boxes.

We should also highlight that we have corrected the interpolation at hub-height for the BSB-lidar cases, resulting in a more accurate representation of the metrics throughout the full manuscript. More specifically, in the first version of the manuscript, we unintentionally used the lidar-measured heights without adding the BSB platform height of 45 m. This led to an incorrect estimation of both wind speed and wind direction at the intended hub height of 107 m.

3.1. General comments

RC: *1. Is there a specific reason for focusing on only one height when the lidars provide velocities at multiple heights? Showing at least one representative vertical profile (e.g.,) would help illustrate the vertical structure of wakes and blockage.*

AR: Dear reviewer, thank you for your comment. We wanted to limit the already large scope and length of the paper, thus, we decided to limit the analysis to hub height, as this is most relevant for operational and planned wind turbines. However, we do agree that an analysis of profiles is interesting, and we would consider conducting it in future studies.

RC: *2. Please comment on the criteria used to select the specific dates/times for the SAR analysis. Were they chosen based on visible wakes, wind direction, stability, or data availability? A short description of the selection process and potential selection bias would be helpful.*

AR: SAR images cannot easily capture wind farm wakes, as they primarily provide wind speeds at approximately 10m height. In addition, only about 8–10 SAR snapshots per month cover the area of interest, and only a few of these contain sufficiently clear wake signatures. Furthermore, we were specifically interested in SW and NE wind directions affecting the Belgian–Dutch cluster, as these can generate wakes that propagate toward the WHi (located SW of the Belgian–Dutch cluster) and EPL (positioned NE the Belgian–Dutch cluster) lidar sites. As a result, the number of suitable SAR snapshots becomes even more limited. For this reason, we carefully selected the snapshots with clearly detectable long wake patterns (rather than short downstream deficits) to evaluate the WRF-WF performance during those cases. The snapshots are illustrative and may therefore reflect selection bias.

The overall text in Section 2.3.2, has been updated as follows (lines 191-201):

The SAR data used in this study were retrieved by the Sentinel-1 European Space Agency mission. More specifically, there are two satellites, Sentinel-1A (2014-present) and Sentinel-1B (2016-2021), equipped with C-band Synthetic Aperture Radar (SAR) instruments operating at 5.405 GHz, and both satellites follow a sun-synchronous polar orbit at an altitude of 693 km [Hasager et al., 2024]. Focusing on the area of interest, including the Belgian-Dutch wind farms cluster in the Southern Bight of the North Sea, as well as additional upstream and downstream areas to study potential wake structures, the analyzed time instances occur approximately at 06:00 UTC every twelve days, and at 17:40 UTC every six days, respectively (Sentinel-1A, 2021-2023). The operation of Sentinel-1B during 2021 provided an additional set of overpasses at around 17:32 UTC every six days. This corresponds to approximately 8 to 10 SAR snapshots per month over the area of interest during the analyzed period. The wind speed at 10-meter height is then estimated by applying the CMOD5N geophysical model function [Hersbach et al., 2007], as explained also in the work of Siedersleben et al. [2020]. This indirect retrieval method assumes that radar backscatter from the sea surface, which reflects centimeter-scale waves generated by instantaneous wind stress, can be converted into wind speed at 10-meter height [Ahsbabs et al., 2017, 2018]. The wind fields are provided at a 500-meter pixel resolution through the Global Wind Atlas Science Portal of the Danish Technical University (DTU) Wind Energy Department (<https://science.globalwindatlas.info/#/map/satwinds> last access: 30 July 2025).

This has also been updated in Section 3.3., lines 504-510

Four specific timestamps have been selected to analyze the wind speed patterns derived from SAR and WRF data, as well as to validate the model performance across the transect that connects the WHi, BSB, and EPL lidars (see Figure 2). These timestamps were selected manually based on the visual appearance of wind farm wakes in the SAR images. Although an average of 8 to 10 snapshots per month (see section 2.3.2) covers the analyzed region, the focus on wake events affecting the WHi, BSB, and EPL lidar sites under SW and NE wind directions resulted in a further reduction of the available snapshots. Therefore, an additional event from September 2020, outside the defined three-year simulation period, has been deliberately included for analysis.

RC: *3. Is the surface roughness used in the SAR-based wind retrieval spatially variable in your analysis? How is this variability calculated or prescribed? A brief explanation of the roughness treatment should be added. In addition, how many SAR snapshots were analyzed in total, and what fraction of them showed clearly visible wakes?*

AR: Only 4 snapshots have been analyzed for this study, and they were manually selected based on visibly long wakes, as described in the previous comment 2.

Regarding the process of retrieving the wind fields from the satellite images, CMOD5N model has been used as explained in the updated Section 2.3.2 (lines 196-199, see comment 2 as well). However, we were not involved in this process, as the DTU database provides directly 10-meter retrieved wind speeds in netcdf format.

The wind speed at 10-meter height is then estimated by applying the CMOD5N geophysical model function [Hersbach et al., 2007], as explained also in the work of Siedersleben et al. [2020]. This indirect retrieval method assumes that radar backscatter from the sea surface, which reflects centimeter-scale waves generated by instantaneous wind stress, can be converted into wind speed at 10-meter height [Ahsbabs et al., 2017, 2018].

RC: *4. Is it possible to quantify the mean wake length for different stability classes (e.g., using a threshold deficit to define “wake end”)? Even a simple diagnostic would strengthen the physical interpretation.*

AR: Since there are already a few studies discussing the wake characteristics (e.g., length) under different stability conditions Ali et al. [2023], Rosencrans et al. [2024], we considered updating our manuscript and assessing the mean wind speed deficit per stability class under the different wind speed regimes. More precisely, Figure 14 has been moved in Section 3.2.3, and it has been updated, including also the average wind speed deficit magnitude per wind speed magnitude class for each ABL stability regime (Figures 14C, and D). Below is the updated text in lines 453-477

Figures 14A, and B present the frequency of the ABL stability conditions under the different wind speed ranges, for SW and NE wind directions, respectively. The ABL stability classification is based on Obukhov length, averaged across the three lidar probes in the NWF simulations. Under low wind speed conditions, very unstable and unstable events tend to dominate. Conversely, a transition toward near-neutral ABL stratification under high wind speeds is also evident. Previous studies [Rosencrans et al., 2024, Palatos-Plexidas et al., 2024, Porchetta et al., 2024] have demonstrated the dependence of wake propagation and wind speed losses on ABL stability, showing that transitioning from neutral to stable and very stable stratification, wakes tend to extend further downstream of wind farms. Similarly, we underscore the relation between wind speed magnitude, ABL stratification, and wake propagation, where during high wind speeds, the frequency of neutral events is enhanced and wakes propagate a few tens of kilometers downstream of the Belgian-Dutch cluster.

In addition to the frequency of the stability regimes, in Figures 14C and D, the magnitude of the average normalized wind speed deficit is presented under the different stability conditions. The wind speed deficits are estimated as the average across the three analyzed lidar locations. It is observed that under very unstable and unstable occurrences, the wake magnitude is relatively lower (varying from 4 to 8%). When ABL stratification is near neutral, higher wake magnitudes are also observed. Following the patterns presented in studies [Rosencrans et al., 2024, Palatos-Plexidas et al., 2024, Porchetta et al., 2024], moving from unstable to neutral and stable stratification results in higher wind speed deficits. However, this should be interpreted together with the frequency of events in each wind speed class, which shows that the number of very stable cases, where wake effects are relatively strong, is not particularly large. On the other hand, the wind speed deficits during very unstable and unstable events that occur more frequently (especially when wind speeds are below 8 m s^{-1}) are significantly lower. Near neutral conditions occur more frequently when wind speed is above 8 m s^{-1} in both SW and NE wind direction, and the wake magnitude in those cases ranges from about 7 to 12%. Greater wind speed magnitudes are associated with more pronounced downstream wind speed deficits (see Figures 12 and 13), as higher wind speeds in this region occur more frequently under neutral to stable conditions. Under these regimes, particularly during stable conditions, reduced turbulent mixing limits wake recovery, leading to enhanced and more persistent wakes [Rosencrans et al., 2024, Ali et al., 2023, Siedersleben et al., 2020]. The transition toward neutral and stable ABL conditions with the increasing wind speed may contribute to the higher cRMSE values observed in Tables 4 and 5, as these regimes are associated with enhanced shear.

In addition, we thought it might be useful to update our manuscript and discuss these characteristics, not under different ABL stratification, but under the different wind speed regimes as presented in Section 3.2.4. Therefore, we have updated Figure 15, including the 95% wake recovery areas (solid purple lines) and the maximum wake length within the area (dashed purple line). The characteristics of the wake are presented in Table 6. Section 3.2.4 has been updated in lines 479-501:

The analysis presented in the previous section, based on SW and NE wind directions, supports the investigation of the two-dimensional wake characteristics and blockage effects within the Belgian-Dutch wind farm cluster. A reduction in wind speed upstream of the wind farms is evident both under SW and NE wind directions. In the transect under SW winds (Figure 12), a noticeable wind speed deficit begins approximately 8 km upstream of the wind farm. This reduction is present across the different wind speed magnitudes, but becomes more rapid around 4 to 5 km upstream, particularly when wind speeds exceed 5 m s^{-1} . For NE wind directions (Figure 13), the blockage effect is also visible, though it is confined to a smaller region, approximately 2 to 3 km upstream of the Borssele wind farm.

The wake maps in Figure 15 further confirm the presence of blockage effects under both SW and NE wind conditions. The solid purple lines indicate the 95% wake recovery area, while the dashed purple lines represent the maximum wake length within this area for each wind speed regime and wind direction. The wake structures become less coherent at wind speeds below 5 m s^{-1} while an induction zone defined by the 95% wake recovery area is observed in low wind speeds. More precisely, the maximum blockage distance from the front row of the wind turbines appears around 10 km under SW wind direction (Figure 15A₁), whereas around 5 km in the NE wind direction regime (Figure 15A₂). In addition, Table 6 presents the values of the 95% wake recovery area and the maximum wake length within this area. The largest coherent wake region is extracted using connected-component labeling. Its perimeter is obtained from the contour of the binary mask, and the maximum extent is computed as the convex-hull diameter of the perimeter points. The estimation of both the 95% wake recovery area and the maximum wake length near SW or NE wind directions is performed using the SciPy Python package[Virtanen et al., 2020]. The values under the different wind speed regimes in Table 6 are comparable for both SW and NE wind directions.

When the flow transitions to higher wind speeds (above 5 m s^{-1}), the wake structures become more coherent, with more pronounced wake recovery areas, leading to overall longer wake lengths (see Table 6). In the $5\text{--}8 \text{ m s}^{-1}$ class, a blockage region forms upstream of the turbines, extending approximately 8 to 9 km from the first row under both SW and NE conditions. At higher wind speeds, this blockage pattern becomes negligible, and the flow tends to deflect around the cluster and propagate farther downstream.

RC: *5. Please discuss the limitations of using the Obukhov length diagnosed by the PBL scheme, particularly given that Monin–Obukhov similarity is known to break down in strongly stable conditions. How might this affect your stability classification and related conclusions?*

AR: There are limitations on the performance of Monin-Obukhov Similarity Theory (MOST), especially under strongly stable conditions. In addition, there are many studies that prefer using other stability metrics like Richardson number. However Obukhov length is still widely used (e.g., Rosencrans et al. [2024], Peña and Hahmann [2012], Quint et al. [2025], Archer et al. [2016]). Therefore, we considered that despite the uncertainties arising from MOST, and the configuration of surface and PBL schemes in WRF, the Obukhov length as a suitable metric to classify stability conditions.

In addition, a short comparison between the Obukhov length L and the bulk Richardson number Ri_b , both obtained from the NWF simulations as WRF outputs (RMOL and BR, respectively), is presented in Figure 1. The two metrics show consistent classification of the three principal stability regimes. The color scale indicates the percentage agreement, and only a small fraction of cases, 1.1% in stable conditions and 2.46% in unstable conditions, do not align. Nevertheless, it is important to note that while L allows a direct and widely used classification based on its sign (i.e., $L < 0$ for unstable, $L > 0$ for stable, and when $L \geq 500$ we consider neutral conditions), the choice of thresholds for Ri_b is less standardized. In this example, we

apply $Ri_b < -0.001$ for unstable conditions, $Ri_b > 0.001$ for stable conditions, and consider the remaining near-zero values as neutral.

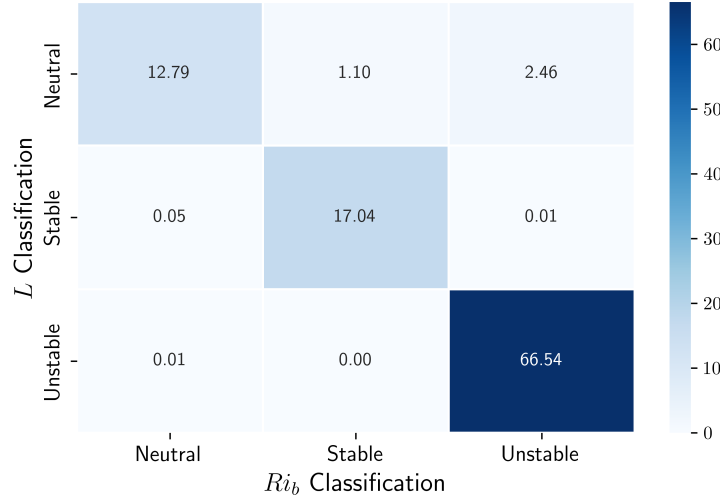


Figure 1: L and Ri_b on classifying stable, neutral, and unstable conditions.

We have updated the revised text in the results Section 3.2.2., as follows in lines 365-369:

The increase in cRMSE under very stable conditions is also reflected in the Pearson correlation, which decreases when transitioning from stable-neutral to very stable regimes. At BSB, all wind speed metrics improve in the WF simulations relative to the NWF runs across the different stability regimes, even under very stable conditions that are known to be associated with enhanced wakes [Rosencrans et al., 2024, Ali et al., 2023, Siedersleben et al., 2020]. These findings align with the correlation analysis between WRF and sonic anemometer measurements in the Horns Rev wind farm [Peña and Hahmann, 2012], which shows that model uncertainty increases as atmospheric stability deviates from neutral, i.e., under very stable or very unstable conditions. *Although the Obukhov length diagnosed by the PBL scheme becomes uncertain under strongly stable conditions when Monin–Obukhov similarity theory is known to break down [Stiperski and Calaf, 2023, Optis et al., 2016] due to weak or intermittent turbulence, it remains a fundamental and practical metric for characterizing ABL stability in mesoscale models. Despite its limitations, many studies use Obukhov length, as it provides a consistent, physically based framework for classifying stability regimes and interpreting wake behaviour [Rosencrans et al., 2024, Quint et al., 2025].*

We consider that the conclusion of the study remains robust, as the Obukhov length still provides the most practical and physically grounded metric for categorizing stability conditions in mesoscale simulations, and the key patterns we report are not sensitive to small shifts in stability class, as the number of very-stable events where the MO similarity theory can break is relatively small.

3.2. Specific comments

RC: *1. Instead of referring only to the lidars by their site names, it may improve readability to also refer to them by their functional roles (e.g., “upwind/free-stream,” “intra-farm,” and “far-wake” or “downstream”)*

throughout the text.

AR: We have improved the manuscript in Sections 3.2.3, 3.2.4, and 3.3, where we discuss the wind farm transects to highlight the positions of the lidars as upstream, intra-farm, and downstream. For example, in the updated Figure 15, there is a distinction of free-stream and down-stream locations. Similarly, in Table 7, which presents the metrics on the SAR events analysis.

RC: **2. Providing the WRF namelist file (at least for the innermost domain) as supplementary material would significantly help reproducibility.**

AR: We will be happy to share our WRF-configuration files upon request. The code availability statement on the manuscript has been updated accordingly.

RC: **3. Figure 7 uses “ r ” for the correlation coefficient, whereas Figures 6 and 8 use “ ρ ” Please standardize the notation across figures for consistency.**

AR: We use the distinction between r and ρ . r is the Pearson correlation coefficient, and it is defined for wind speed (and linear variables in general), while ρ is the circular (directional) correlation coefficient. Although both estimate the correlation between the observations and the simulations, their definitions differ. The equations are provided in the Appendix, and lines 238-239 have been added in Section 2.3.4.

To compute means on the wind direction, the circular means are applied [Mardia and Jupp, 2009]. The cRMSE equation for circular variables (A6) has been adopted from the study of [Rosencrans et al., 2024]. *Instead of the Pearson correlation for linear values, the circular-circular correlation ρ is used [Mardia and Jupp, 2009].* All metrics are computed using the built-in functions of the NumPy [Harris et al., 2020] and SciPy [Virtanen et al., 2020] Python packages. Furthermore, the *wasserstein_circle* function by the Python Optimal Transport [Flamary et al., 2021] package is used to calculate the EMD for the wind direction.

References

- T. Ahsbahs, M. Badger, P. Volker, K. S. Hansen, and C. B. Hasager. Applications of satellite winds for the offshore wind farm site Anholt. *Wind Energy Science*, 3(2):573–588, 2018. . URL <https://wes.copernicus.org/articles/3/573/2018/>.
- Tobias Ahsbahs, Merete Badger, Ioanna Karagali, and Xiaoli Guo Larsén. Validation of sentinel-1a sar coastal wind speeds against scanning lidar. *Remote Sensing*, 9(6), 2017. ISSN 2072-4292. . URL <https://www.mdpi.com/2072-4292/9/6/552>.
- Tobias Ahsbahs, Nicolai Gayle Nygaard, Alexander Newcombe, and Merete Badger. Wind farm wakes from sar and doppler radar. *Remote Sensing*, 12(3), 2020. ISSN 2072-4292. . URL <https://www.mdpi.com/2072-4292/12/3/462>.
- Karim Ali, David M Schultz, Alistair Revell, Timothy Stallard, and Pablo Ouro. Assessment of five wind-farm parameterizations in the weather research and forecasting model: a case study of wind farms in the north sea. *Monthly Weather Review*, 151(9):2333–2359, 2023. URL <https://doi.org/10.1175/MWR-D-23-0006.1>.
- Cristina L Archer, Brian A Colle, Dana L Veron, Fabrice Veron, and Matthew J Sienkiewicz. On the predominance of unstable atmospheric conditions in the marine boundary layer offshore of the us

- northeastern coast. *Journal of Geophysical Research: Atmospheres*, 121(15):8869–8885, 2016. URL <https://doi.org/10.1002/2016JD024896>.
- Cristina L Archer, Sicheng Wu, Yulong Ma, and Pedro A Jiménez. Two corrections for turbulent kinetic energy generated by wind farms in the wrf model. *Monthly Weather Review*, 148(12):4823–4835, 2020. URL <https://doi.org/10.1175/MWR-D-20-0097.1>.
- R Borgers, J Meyers, and N P M van Lipzig. Energy production and inter-farm wake losses in future north sea wind farms. *Environmental Research Letters*, 20(7):074036, jun 2025. . URL <https://doi.org/10.1088/1748-9326/add8a2>.
- B. Cañadillas, M. Beckenbauer, J. J. Trujillo, M. Dörenkämper, R. Foreman, T. Neumann, and A. Lampert. Offshore wind farm cluster wakes as observed by long-range-scanning wind lidar measurements and mesoscale modeling. *Wind Energy Science*, 7(3):1241–1262, 2022. . URL <https://wes.copernicus.org/articles/7/1241/2022/>.
- Beatriz Cañadillas, Richard Foreman, Volker Barth, Simon Siedersleben, Astrid Lampert, Andreas Platis, Bughsin Djath, Johannes Schulz-Stellenfleth, Jens Bange, Stefan Emeis, and Thomas Neumann. Offshore wind farm wake recovery: Airborne measurements and its representation in engineering models. *Wind Energy*, 23(5):1249–1265, 2020. . URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/we.2484>.
- L. de Montera, T. Remmers, R. O’Connell, and C. Desmond. Validation of sentinel-1 offshore winds and average wind power estimation around ireland. *Wind Energy Science*, 5(3):1023–1036, 2020. . URL <https://wes.copernicus.org/articles/5/1023/2020/>.
- J. Fischereit, H. Vedel, X. G. Larsén, N. E. Theeuwes, G. Giebel, and E. Kaas. Modelling wind farm effects in harmonie–arome (cycle 43.2.2) – part 1: Implementation and evaluation. *Geoscientific Model Development*, 17(7):2855–2875, 2024a. . URL <https://gmd.copernicus.org/articles/17/2855/2024/>.
- J. Fischereit, H. Vedel, X. G. Larsén, N. E. Theeuwes, G. Giebel, and E. Kaas. Modelling wind farm effects in harmonie–arome (cycle 43.2.2) – part 1: Implementation and evaluation. *Geoscientific Model Development*, 17(7):2855–2875, 2024b. . URL <https://gmd.copernicus.org/articles/17/2855/2024/>.
- Jana Fischereit, Roy Brown, Xiaoli Guo Larsén, Jake Badger, and Graham Hawkes. Review of mesoscale wind-farm parametrizations and their applications. *Boundary-Layer Meteorology*, 182(2):175–224, 2022. URL <https://doi.org/10.1007/s10546-021-00652-y>.
- Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z Alaya, Aurélie Boissunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, et al. Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8, 2021. URL <http://jmlr.org/papers/v22/20-451.html>.
- O. García-Santiago, A. N. Hahmann, J. Badger, and A. Peña. Evaluation of wind farm parameterizations in the wrf model under different atmospheric stability conditions with high-resolution wake simulations. *Wind Energy Science*, 9(4):963–979, 2024. . URL <https://wes.copernicus.org/articles/9/963/2024/>.

- Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. . URL <https://doi.org/10.1038/s41586-020-2649-2>.
- Charlotte B. Hasager, Merete Badger, Alfredo Peña, Xiaoli G. Larsén, and Ferhat Bingöl. Sar-based wind resource statistics in the baltic sea. *Remote Sensing*, 3(1):117–144, 2011a. ISSN 2072-4292. . URL <https://www.mdpi.com/2072-4292/3/1/117>.
- Charlotte B. Hasager, Merete Badger, Alfredo Peña, Xiaoli G. Larsén, and Ferhat Bingöl. Sar-based wind resource statistics in the baltic sea. *Remote Sensing*, 3(1):117–144, 2011b. ISSN 2072-4292. . URL <https://www.mdpi.com/2072-4292/3/1/117>.
- Charlotte Bay Hasager, Pauline Vincent, Jake Badger, Merete Badger, Alessandro Di Bella, Alfredo Peña, Romain Husson, and Patrick J. H. Volker. Using satellite sar to characterize the wind flow around offshore wind farms. *Energies*, 8(6):5413–5439, 2015. ISSN 1996-1073. . URL <https://www.mdpi.com/1996-1073/8/6/5413>.
- Charlotte Bay Hasager, James Imber, Jana Fischereit, Aito Fujita, Krystallia Dimitriadou, and Merete Badger. Wind speed-up in wind farm wakes quantified from satellite sar and mesoscale modeling. *Wind Energy*, 27(11):1369–1387, 2024. URL <https://doi.org/10.1002/we.2943>.
- H. Hersbach, A. Stoffelen, and S. de Haan. An improved c-band scatterometer ocean geophysical model function: Cmod5. *Journal of Geophysical Research: Oceans*, 112(C3), 2007. . URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006JC003743>.
- X. G. Larsén and J. Fischereit. A case study of wind farm effects using two wake parameterizations in the weather research and forecasting (wrf) model (v3.7.1) in the presence of low-level jets. *Geoscientific Model Development*, 14(6):3141–3158, 2021. . URL <https://gmd.copernicus.org/articles/14/3141/2021/>.
- Kanti V Mardia and Peter E Jupp. *Directional statistics*. John Wiley & Sons, 2009. .
- B. T. E. Olsen, A. N. Hahmann, N. G. Alonso-de Linaje, M. Žagar, and M. Dörenkämper. Low-level jets in the north and baltic seas: mesoscale model sensitivity and climatology using wrf v4.2.1. *Geoscientific Model Development*, 18(14):4499–4533, 2025. . URL <https://gmd.copernicus.org/articles/18/4499/2025/>.
- Michael Optis, Adam Monahan, and Fred C. Bosveld. Limitations and breakdown of monin–obukhov similarity theory for wind profile extrapolation under stable stratification. *Wind Energy*, 19(6):1053–1072, 2016. . URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/we.1883>.
- Alexandros Palatos-Plexidas, Simone Gremmo, Sara Porchetta, Jeroen Van Beeck, Lesley De Cruz, and Wim Munters. A numerical analysis of wind farm wake characteristics in the southern part of the north sea. *Journal of Physics: Conference Series*, 2767(9):092078, jun 2024. . URL <https://dx.doi.org/10.1088/1742-6596/2767/9/092078>.
- Alfredo Peña and Andrea N. Hahmann. Atmospheric stability and turbulence fluxes at horns rev—an intercomparison of sonic, bulk and wrf model data. *Wind Energy*, 15(5):717–731, 2012. . URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/we.500>.

- Andreas Platis, Simon K. Siedersleben, Jens Bange, Astrid Lampert, Konrad Bärfuss, Rudolf Hankers, Beatriz Cañadillas, Richard Foreman, Johannes Schulz-Stellenfleth, Bughsin Djath, Thomas Neumann, and Stefan Emeis. First in situ evidence of wakes in the far field behind offshore wind farms. *Scientific Reports*, 8(1): 2163, 2018. ISSN 2045-2322. . URL <https://doi.org/10.1038/s41598-018-20389-y>.
- S. Porchetta, M. F. Howland, R. Borgers, S. Buckingham, and W. Munters. Annual wake impacts in and between wind farm clusters modelled by a mesoscale numerical weather prediction model and fast-running engineering models. *Wind Energy Science Discussions*, 2024:1–37, 2024. . URL <https://wes.copernicus.org/preprints/wes-2024-58/>.
- V. Pronk, N. Bodini, M. Optis, J. K. Lundquist, P. Moriarty, C. Draxl, A. Purkayastha, and E. Young. Can reanalysis products outperform mesoscale numerical weather prediction models in modeling the wind resource in simple terrain? *Wind Energy Science*, 7(2):487–504, 2022. . URL <https://wes.copernicus.org/articles/7/487/2022/>.
- D. Quint, J. K. Lundquist, and D. Rosencrans. Simulations suggest offshore wind farms modify low-level jets. *Wind Energy Science*, 10(1):117–142, 2025. . URL <https://wes.copernicus.org/articles/10/117/2025/>.
- D. Rosencrans, J. K. Lundquist, M. Optis, A. Rybchuk, N. Bodini, and M. Rossol. Seasonal variability of wake impacts on us mid-atlantic offshore wind plant power production. *Wind Energy Science*, 9(3): 555–583, 2024. . URL <https://wes.copernicus.org/articles/9/555/2024/>.
- Ameya Sathe, Sven-Erik Gryning, and Alfredo Peña. Comparison of the atmospheric stability and wind profiles at two wind farm sites over a long marine fetch in the north sea. *Wind Energy*, 14(6):767–780, 2011. . URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/we.456>.
- S. K. Siedersleben, A. Platis, J. K. Lundquist, B. Djath, A. Lampert, K. Bärfuss, B. Cañadillas, J. Schulz-Stellenfleth, J. Bange, T. Neumann, and S. Emeis. Turbulent kinetic energy over large offshore wind farms observed and simulated by the mesoscale model wrf (3.8.1). *Geoscientific Model Development*, 13(1): 249–268, 2020. . URL <https://gmd.copernicus.org/articles/13/249/2020/>.
- Ivana Stiperski and Marc Calaf. Generalizing monin-obukhov similarity theory (1954) for complex atmospheric turbulence. *Phys. Rev. Lett.*, 130:124001, Mar 2023. . URL <https://link.aps.org/doi/10.1103/PhysRevLett.130.124001>.
- Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272, 2020. URL <https://doi.org/10.1038/s41592-019-0686-2>.
- P. J. H. Volker, J. Badger, A. N. Hahmann, and S. Ott. The explicit wake parametrisation v1.0: a wind farm parametrisation in the mesoscale model wrf. *Geoscientific Model Development*, 8(11):3715–3731, 2015. . URL <https://gmd.copernicus.org/articles/8/3715/2015/>.
- L. Vollmer, B. A. M. Sengers, and M. Dörenkämper. Brief communication: A simple axial induction modification to the weather research and forecasting fitch wind farm parameterization. *Wind Energy Science*, 9(8):1689–1693, 2024. . URL <https://wes.copernicus.org/articles/9/1689/2024/>.