

# Revision updates for WES-2023-33

September 15, 2023

## Note to the editor

Since the initial submission of the manuscript, some additional points of improvement were identified next to the suggestions of the referees. Most of these points are minor, however, as already discussed with the associate editor, one major point of improvement was identified. We thank the associate editor for the patience and the supportive attitude in dealing with this major point. Therefore, this document contains three sections: first the updates are described that were done and were not related to the referee comments (section 1). This is then followed by a list of comments and replies to the first referee (section 2) and then finally the same for all the comments of the second referee (section 3).

## 1. Additional updates after the initial submission

**Point 1:** A code bug led to the modelled thrust being too low in the WFP implementation compared to the initial description of Fitch (2012).

**Update:** Testing indicated significant differences with the initial simulations. Therefore, the four wind farm simulations were done again with the code bug fixed. As a consequence, the results that were mentioned in the abstract, results and discussion and conclusion sections have been adapted. The general change is that wind farm wake effects are now stronger and all signals have been amplified to some extent. The text body, Figures 7 (bottom panel), 8, 9 and 10 and Table 4 have now been adapted with the new data. Most conclusions have remained the same, albeit with stronger magnitudes. However, in some locations the text body, sentences no longer valid have been removed and have been replaced. This is visualized in detail in the provided latexdiff document (mainly section 3.2). At the bottom of the manuscript, a code availability statement was added, referring to the code and the implementation guide (which includes an update log of all the updates compared to the previous version). The new WFP implementation in COSMO5.0 has been added as a next version to the repository of the initial code: <https://www.wdc-climate.de/ui/entry?acronym=WindFarmPCOSMO5.0clm15>

**Point 2:** Simulation and evaluation domain depicted incorrectly in Fig. 1 due to inadequate .csv description of the domain.

**Update:** A better domain description was generated and the outlines of the domains have been adapted accordingly in Fig. 1.

**Point 3:** The amount of grid cells that were clipped at the edges to remove the relaxation and spin-up zones was too low.

**Update:** An extra 10 grid cells were now clipped at the edges. This has slightly changed the results of Fig. 7 and table 4. For Table 4, this mainly pertains to columns 3 and 6. In column 3, the total capacity in the evaluation domain is slightly reduced and in column 6, the layout-integrated AEP has reduced accordingly. Importantly, results in column 6 are also influenced by the changes related to point 1 of this document. The impact on all other results, which are mainly efficiency indicators, is negligible.

**Point 4:** The computation of the the capacity factors for the full domain per grid cell (only Fig. 7 and table 4) had a small error in the python code leading to slightly too low capacity factors.

**Update:** The code was fixed leading to slightly higher capacity factors (a few percents). It was checked that indeed this error was not present in the computation of all other results.

**Point 5:** Acknowledgements needed to be added for co-author Naveed Akhtar.

**Update:** One acknowledgement has now been added.

## 2. Comments and updates for referee 1: David Schultz

### General feedback from the referee

This is an important paper because it shows that the increase in height of the turbines from 5 MW to 15 MW more than offsets the loss due to wakes. I approach this review as a meteorologist and mesoscale modeler. Hence, my comments are mostly focused on these aspects of the manuscript. I see no problem with the science and the results, only the presentation.

### General reply

The authors thank the referee for the thorough assessment of the text and the suggested adaptations to improve the presentation of our results and the text body. We believe that after having processed these comments, the presentation and text are now clearer and more streamlined across the manuscript which is very much appreciated. We summarize here all the updates done to the initial submission based on the different comments of the reviewer.

### Major comments

**Comment 1:** The conclusion section is unbalanced. Lines 414–423 represent an unacceptably short, incomplete, and qualitative summary of the first part of the study. In contrast, lines 424–443 represent a much more detailed and quantitative set of conclusions. I prefer the latter, as I imagine most readers would who would want to read the conclusion to get a more clear picture of the results of this study. I suggest a revision to the first part of the conclusion section.

**Update:** The evaluation part of the conclusion section has now been elaborated: more details are provided with a focus on more quantitative information. This has made the conclusion more balanced between the two parts of the study and we thank the reviewer for pointing this out. The conclusion of the evaluation part is now:

*First, the model skill in simulating the wind climate was evaluated in a comparison with in situ, lidar and satellite data, which revealed that*

- *The differences between the measured and modelled, long-term mean wind speed at turbine hub height ( $\sim 100$  m) are generally within the measurement uncertainty. This is also the case for differences at higher altitudes (up to 290 m), but closer to the surface COSMO-CLM underestimates the mean wind speed ( $\sim -0.5$  ms<sup>-1</sup>). Under stable stratification ( $\sim 25\%$ ), the model underestimates the long term mean wind speed at turbine height, but not under weakly stable and unstable stratification ( $\sim 75\%$ ).*
- *The agreement between the measured and modelled, long-term wind speed histograms is high, with a PSS above 95% in most cases. The theoretical capacity factors derived from these histograms agree well overall, but small underestimations ( $\sim 1-5\%$ ) are present at some locations.*
- *The agreement with the wind speed measurements is consistent over the different years of the simulation period as inter-annual variations in the mean wind speed difference and the PSS are small. Also the seasonal variability in the shape and location of the wind speed distribution is captured by COSMO-CLM.*
- *Multi-year histograms of wind direction also agree well, with again a PSS above 95% in most cases. Also the variation of the wind speed histograms over 12 directional bins ( $30^\circ$ ) are adequately captured in the model. This encourages the application of COSMO-CLM to wind farm modelling as wind farm shapes are adapted to the regional wind climate.*

*As deviations mainly occur under stable conditions, a stability-dependent bias correction could be considered for future applications in addition to continuous efforts to improve the model. Overall, this evaluation emphasizes the value of having a large set of wind measurements available in regions for offshore wind farm development, as it allows a benchmarking of mesoscale models over the region of interest.*

**Comment 2:** Figure 2: The color scale needs work. First, the zero point should be white, not bluish-yellow, to indicate its true neutrality. It's hard to interpret otherwise. Second, the color scheme is not symmetric. Negative

values are all shades of blue, but positive values are yellow, orange, and red. Instead, all positive values should be shades of red, opposite of the negative values. In any case, I recommend to the authors to avoid the rainbow color scheme. It distorts gradients, among other issues. Please fix similar problems with Figures 3 and S5.

**Update:** The color scale of Fig. 2, Fig. 3 and Fig. S5 (now Fig. S6 in the revision) has been replaced with a blue-white-red scale, which is more symmetric and has white as the zero point. While still considering colorblindness, this color scale now makes the figures easier to interpret. The new figures are included in the revised manuscript.

**Comment 3:** An excessive number of grid lines appears on Figures 5, 6, S1–S4, and S6–S9. These are distracting from the data (i.e., what Edward Tufte calls chartjunk) and should be eliminated.

**Update:** The number of gridlines has been minimized for Fig. 5, Fig. 6, Fig. S1, S2, S3 and S6 (now Fig. S7 in the revision). For Fig. S4, S7–S9, the gridlines have been kept the same since we believe it is easier to interpret the figures with these present. The figures are included in the revised manuscript.

**Comment 4:** Model output and observations have different colors in each figure where they are compared directly against each other: Figures 4, 5, and 6 (and comparable figures in the supplement). Can a similar color scheme (red for model and black for observations, for example) be employed across all three of these figures? It would sure simplify things for the reader having that consistency across all the figures.

**Update:** Figures 4, 5 and the bottom part of figure 6 have now all the combination: grey for the observations - orange for the model. This was found to be the best choice for streamlining the color combination across the relevant figures, also considering colorblindness. The figures are included in the revised manuscript.

**Comment 5:** Lines 282–284: I find three levels of organization (section 3 to subsection 3.1 to subsection 3.1.1) without any text introducing each confusing. The authors need to put these sections into context before diving down three levels to a very specific quantity (e.g., wind speed at 290 m). For example, after the title for section 3, there should at least be a paragraph explaining how this section is structured and any general information that pertains to it. Also, after the title for section 3.1, there should at least be a paragraph explaining what will be discussed in this subsection and any general information that pertains to it. The same is true for section 3.2. The text just jumps right in with results from Figure 7. Can you provide some context to the reader first?

**Update:** At the start of the two sub-sections a short, situating description has been added for the reader. For the evaluation section 3.1:

*This subsection covers the model performance evaluation. First, the general evaluation based on all validation sources and the complete height range (10 m to 290 m) is described. This is followed by a more detailed performance analysis at turbine hub-height (~100 m) and finally the evaluation is extended to the different atmospheric stability classes.*

And for the wind farm simulation intercomparison, i.e. section 3.2:

*This subsection covers the results of the wind farm simulations. First, the impact of the NREL8.1 base scenario on the wind climate and wind resource is described, also under different atmospheric stability conditions. Afterwards, the different wind farm scenarios are compared in terms of cluster-scale wake effects and efficiency of power production.*

**Comment 6:** Line 284: Are three levels of organization necessary? Usually sections and subsections are sufficient. Could you just have different sections for results? Section 3.1 becomes section 3, section 3.2 becomes section 4, etc.? Alternatively, you could just drop the third level of subsections. I'm not sure they are helping the reader. It would make the text more readable and less tedious. Also, the text of section 3 is just a little over 100 lines. Three levels of subsections seems excessive.

**Update:** The sub-subsection organization of the evaluation part, i.e. 3.1.1, 3.1.2 and 3.1.3 have now been removed so that all these parts are merged to sub-section "3.1 evaluation", hence only two levels of organization remain here.

## Minor comments

**Comment 1:** Lines 13–15: “In contrast, the impact of wake losses is exacerbated with increasing capacity density, as the layout-integrated, annual capacity factor varies between 54.4% and 44.3% over the considered range of 3.5 to 10 MW km<sup>2</sup>.” I have read this sentence multiple times, and I am having trouble making sense of it. I think it is phrase “impact of wake losses is exacerbated” that is the problem. Could it be written more clearly?

**Update:** This has been adapted to: “In contrast, the efficiency losses due to wakes become larger with increasing capacity density as the layout-integrated, annual capacity factor varies between 51.8% and 38.2% over the considered range of 3.5 to 10 MW km<sup>-2</sup>”

*In contrast, the efficiency losses due to wakes become larger with increasing capacity density as the layout-integrated, annual capacity factor varies between 51.8% and 38.2% over the considered range of 3.5 to 10 MW km<sup>-2</sup> - this is indeed clearer.*

**Comment 2:** Lines 15–16: “wind farm characteristics and inter-farm distances play an essential role in cluster-scale wake losses, which should be taken into account in future wind farm planning.” This sentence is too vague and general to be a meaningful conclusion to your paper. For example, this sentence could be the conclusion of just about every paper in Wind Energy Science. It does not do your study justice.

**Update:** We improved the conclusive statement to:

*In conclusion, our results show that the wake losses in future wind farm clusters are highly sensitive to the inter-farm distances and the capacity densities of the individual wind farms and that the evolution of turbine technology plays a crucial role in offsetting these wake losses.*

**Comment 3:** Lines 25, 33: Why these three citations of all the citations that have been published on the efficiency of wind-farm wakes? In such instances, it would help to precede the list of citations with “e.g.” to indicate that these are a sampling of all the possible sources that could have been cited. You may wish to consider adding “e.g.” to other lists of citations, as well.

**Update:** An “e.g.” has been added to all the citations in the text where this comment is relevant

**Comment 4:** Line 64: In contrast, the “e.g.” is not needed here because presumably there is only source for “documentation”.

**Update:** The “e.g.” is retained because it refers to only one of the chapters of the model documentation.

**Comment 5:** Line 47: should be “and/or”, with no spaces

**Update:** Changed as suggested.

**Comment 6:** Line 67: The verb tense changes back and forth from present to past tense in this paragraph. I think past tense sounds better, but whatever you pick, aim to be consistent.

**Update:** This has been adapted to the past tense throughout the paragraph except where another tense (e.g. present) is really needed.

**Comment 7:** Line 71: I think more careful wording of “deep convection is explicitly resolved at the mesogamma scale” is needed. Convective storms may start to be resolved at these scales, but the updrafts comprising that convection are not, as has been shown in Bryan et al. (2003). So, the phrase “deep convection is explicitly resolved” is ambiguous.

**Update:** This has been adapted to: "At the meso- $\gamma$  scale, the model resolution partly allows the explicit development of deep convection so that only shallow convection was parametrized according to the scheme of Tiedtke (citation). Switching of the deep convection parametrization on this resolution has previously been shown not to degrade COSMO simulations (Vergara-Temprado et al., 2020).

**Comment 8:** Lines 80, 340: “Hence” cannot be used as a conjunction in this context.

**Update:** Updated for both cases.

**Comment 9:** Line 89: I’m unclear what role these two citations are supporting. Period of 13 years? Need for sampling inter-annual variability? Other studies doing model evaluation? More explanation is needed for why their citation is relevant to this sentence. That may involve rewriting the sentence so that the reason for the citations becomes more clear.

**Update:** This sentence has been rewritten to make this more clear. The adapted version is: "To evaluate the model performance, a simulation was performed for a period of 13 years (2008–2020). Data from in situ, lidar and satellite measurements over the North Sea are abundant in both space and time for this period. Additionally, the length of the simulation ensures that the inter-annual variability in the wind conditions, which has been widely described (e.g. Geyer et al., 2015; Ronda et al., 2017), is sampled well."

**Comment 10:** Lines 93, 264, 367, 396: “Cf.” means “compare”. So what is being compared to section 2.4.1 It isn’t clear. Write instead “cf. A and B” to compare A to B.

**Update:** We recognized that "cf." was used incorrectly in our text. All instances in the text have been changed to not include this.

**Comment 11:** Line 99: Change “constructions” to “construction”.

**Update:** Adapted accordingly.

**Comment 12:** Line 104: Delete the comma.

**Update:** Adapted accordingly.

**Comment 13:** Lines 107, 112, 213, 216, 243, 269, 402, 424: Please insert noindent before “where”.

**Update:** All instances in the text have been adapted with noindent.

**Comment 14:** Lines 104 and elsewhere: Italicize  $R$ , as it is a variable

**Update:** All variables in the manuscript, such as  $R$  but also  $PSS$  have now been italicized in the text body.

**Comment 15:** Line 112: Italicize b.

**Update:** Has now been italicized.

**Comment 16:** Lines 109, 116, and elsewhere: Change hyphens to en dashes (i.e., two hyphens in LaTeX) to connect two items in a range. Fix also in Tables A1 and A2. Fix throughout the manuscript.

**Update:** All instances throughout the text where this update was needed have been adapted to en dashes.

**Comment 17:** Line 118: Capitalize “Figures”

**Update:** This has been capitalized now.

**Comment 18:** Lines 118–120: I am having a hard time understanding this sentence. I think there is too much being communicated within. Break it up, perhaps.

**Update:** This sentence has now been broken up as suggested. The new version is: “Based on this procedure, the year 2016 was selected for the simulations, as the representativeness is high overall for this year (supplementary Fig. S1). In addition, the representativeness is especially high for wind direction (supplementary Fig. S2), which is particularly important for the study of inter-farm wake interactions.”

**Comment 19:** Lines 118–120: On what basis can it be said that “the representativeness is high overall, but especially for the wind direction distribution,”? Please evidence that statement.

**Update:** References to supplementary figures have now been added to these statements. These were previously only referred to in the previous sentence, but have now been included explicitly for these statements.

**Comment 20:** Lines 157, 219, 230: Is this a new paragraph? If so, indent it. If not, then combine it with the previous line.

**Update:** For each instance this has now been evaluated and corrected if indeed it was still belonging to the previous paragraph.

**Comment 21:** Line 164: Change “since” to ”because” to avoid implication of time that “since” implies. See also lines 207, 239, 257, 266.

**Update:** This has been adapted to ”because” for all the instances in the text where the implication of time was indeed wrong.

**Comment 22:** Line 170: spell out “including”

**Update:** This has been done as proposed.

**Comment 23:** Lines 195, 311, 445, 447: Change “while” to ”although” to avoid the implication of simultaneity that “while” implies.

**Update:** This has been adapted to ”although” for all the instances in the text where the implication of time was indeed wrong.

**Comment 24:** Line 199: Change “resolution” to “grid spacing” for consistency and precision

**Update:** This has been done as proposed.

**Comment 25:** Line 208 and throughout the manuscript: Hyphenate “10-minute period”. Fix throughout.

**Update:** This has been hyphenated for all instances in the text.

**Comment 26:** Lines 214, 241, 267: Delete the colon

**Update:** These have been removed.

**Comment 27:** Line 270: Italicize P.

**Update:** This has been done.

**Comment 28:** Line 295: “further” should be “farther”.

**Update:** This has been adapted.

**Comment 29:** Line 297–299: This sentence should cite Fig. 3. You were talking about Figure 3, but then cited Figure 2. You should return to citing Figure 3 to avoid any confusion and point the reader to the right figure.

**Update:** The figure referencing here was indeed not clear enough. The references per sentence have now been updated to fix this, now including the correct use of “cf.” ;). The new version is: “The model underestimation near the surface that was identified against the in situ data in the southern North Sea is much smaller than the differences compared to ASCAT in this region (cf. Fig. 2 and Fig. 3). A three-way comparison with three in situ stations shows that the mean differences against the in situ data exceed the in situ measurement uncertainty for both COSMO-CLM and ASCAT (Fig. 3).”

**Comment 30:** Figure 3: If the top part of the graph is gray (i.e., presumably no difference field plotted there), then why not trim it off? Also, does it make sense to make the graph encompass the same domain as Figure 7? If so, that would be helpful to the readers to compare figures that have the exact same map background.

**Update:** This map has been adapted in such a way as to balance the two suggestions of the referee: The northern gray zone has been clipped, but at the same time matched to the maps of Fig. 7 i.e. the actual extent of the evaluation domain.

**Comment 31:** Figure 4: I am having a hard time understanding this graph. There are three colors (light green, dark green, and gray), yet only two colors are explained in the figure caption. Can you write the figure caption more clearly

(or replot the graph) to make it easier to understand? I think the histograms are overlapping, but this is difficult to interpret.

**Update:** The caption of this figure has been updated to explain the three different colors visible in this plot.

**Comment 32:** Line 315: Is “stable” the right word? “Consistent” would be better.

**Update:** This has indeed been updated to “consistent”.

**Comment 33:** Line 343: The title of this subsection is vague: “characteristics”. It could be more clear what characteristics are being varied.

**Update:** This title has been updated from “Effect of wind farm characteristics” to: “Impact of wind farm characteristics on cluster-scale wake losses”. The title is now more concrete and less vague, however the characteristics have not been written out explicitly because, as described in the methods section, many factors are investigated: turbine type, capacity density, inter-farm spacing and wind farm size. The title would be too long if we include all of these, but we believe that the abstract, introduction, methods, conclusion and the results section itself make clear enough what characteristics we investigate.

**Comment 34:** Line 347: Put a comma between “conditions capacity”.

**Update:** This has been added.

**Comment 35:** Line 360: Change “while” to “whereas” to avoid implication of simultaneity.

**Update:** Changed accordingly.

**Comment 36:** Line 364: Why is “weakly” in parentheses? You don’t include the parentheses when that term is used in line 366. I suggest deleting the parentheses.

**Update:** This was to refer to both the weakly stable and stable classes simultaneously, but was actually a poor choice. This has now been adapted to “stable and weakly stable”.

**Comment 37:** Figure 8 caption: Use the  $\pm$  symbol.

**Update:** This has now been used instead.

**Comment 38:** Lines 373, 384: Delete “(blue)” as that information belongs in the legend and figure caption, not the text.

**Update:** This has been removed.

**Comment 39:** Line 387: Change “wit” to “with”.

**Update:** This has been corrected.



**Comment 40:** Line 404: Be careful of the difference between cite and citep.

**Update:** These citations have been corrected.

**Comment 41:** Line 404: Insert a comma after “IEA8.1”.

**Update:** comma has been inserted.

**Comment 42:** Line 405: Insert a comma after “IEA10.0”.

**Update:** comma has been inserted.

**Comment 43:** Lines 414–415: “negative bias” in “wind climate” is unclear. Can you just say that the “model was underpredicting the wind speed”? That is easier to understand. Look for other similar opportunities throughout the manuscript to convey your message more simply and more clearly.

**Update:** This piece of text has been updated and, as suggested, we have gone through the evaluation parts of the study to convey the message in a more easily understandable way. This is noticeable in the latexdiff document.

**Comment 44:** Line 417: A comma is needed after “95%” to join the two independent clauses

**Update:** This comma has been added

**Comment 45:** Line 417: Delete “both” because you describe three things: “seasonal, inter-annual, and directional variability”.

**Update:** This has indeed been removed.

**Comment 46:** Lines 418–419: This sentence is unclear. Also, what about biases in other stability conditions?

**Update:** The sentence “Separation into stability classes reveals small, height-dependent biases under dynamically stable conditions” has been updated within the updated conclusion to: “... Under stable stratification (~25%), the model underestimates the long term mean wind speed at turbine height, but not under weakly stable and unstable stratification (~75%).”

### 3. Comments and updates for referee 2: Andrea Hahmann

#### General feedback from the referee

The manuscript presents an excellent contribution to assessing wind resources in the North Sea, which could be limited by extracting kinetic energy from the atmosphere by large wind farms offshore. The work is well embedded in the existing literature and brings enough novelty. The design of the study is robust, and results soundly support the

conclusions, and I recommend the publication with two ‘medium’ and a few minor concerns, as follows:

## General reply from the authors

The authors thank the referee again for taking the time to go through our work and formulate some very insightful suggestions to help us in making our work more robust and clear. Here we summarize the different updates that have been done to the initial submission based on the different comments of the referee.

## Medium comments

**Comment 1:** It will be nice to get an indication of the accuracy of the simulated model stability classes. Since you are using these frequencies to split the wake losses among stability classes, it would be good to know if they relate to reality. We know the SSTs are input to the model simulation, so could we verify the temperature above the sea using buoy data? For the NEWA project, we estimated that the stability classes could be different by as much as 10% when a different PBL was used in the simulations (see Figures 11-13 in NEWA D4.3 report).

**Update:** As we proposed in the reply, we have handled this concern by performing a temperature gradient validation at Meteomast Ijmuiden. We believe that a validation over the available height range (21m-90m) is more suitable than close to the surface, because it is closer to the height range over which our stability criterion is evaluated. This validation was done for a 4-year period (2012–2015). An additional supplementary figure (supplementary Fig. S5) has now been added which shows the long-term histograms of the temperature gradient for the measurements at the station and the corresponding model data. Although the match is not perfect, the agreement is good (along with a negligible difference in the mean) which supports the validity of doing the stability subdivision with the Bulk Richardson criterion as we did it. At line 257 in the old manuscript the following text has been added to discuss this comparison:

*A comparison of the modelled temperature gradients with measured temperature gradients at the station MMIJ between 90 m MSL and 21 m MSL shows a good correspondence in the long-term temperature gradient probability distribution, indicating sufficient model skill for this subdivision into stability classes (supplementary Fig. S5).*

And the added figure is:

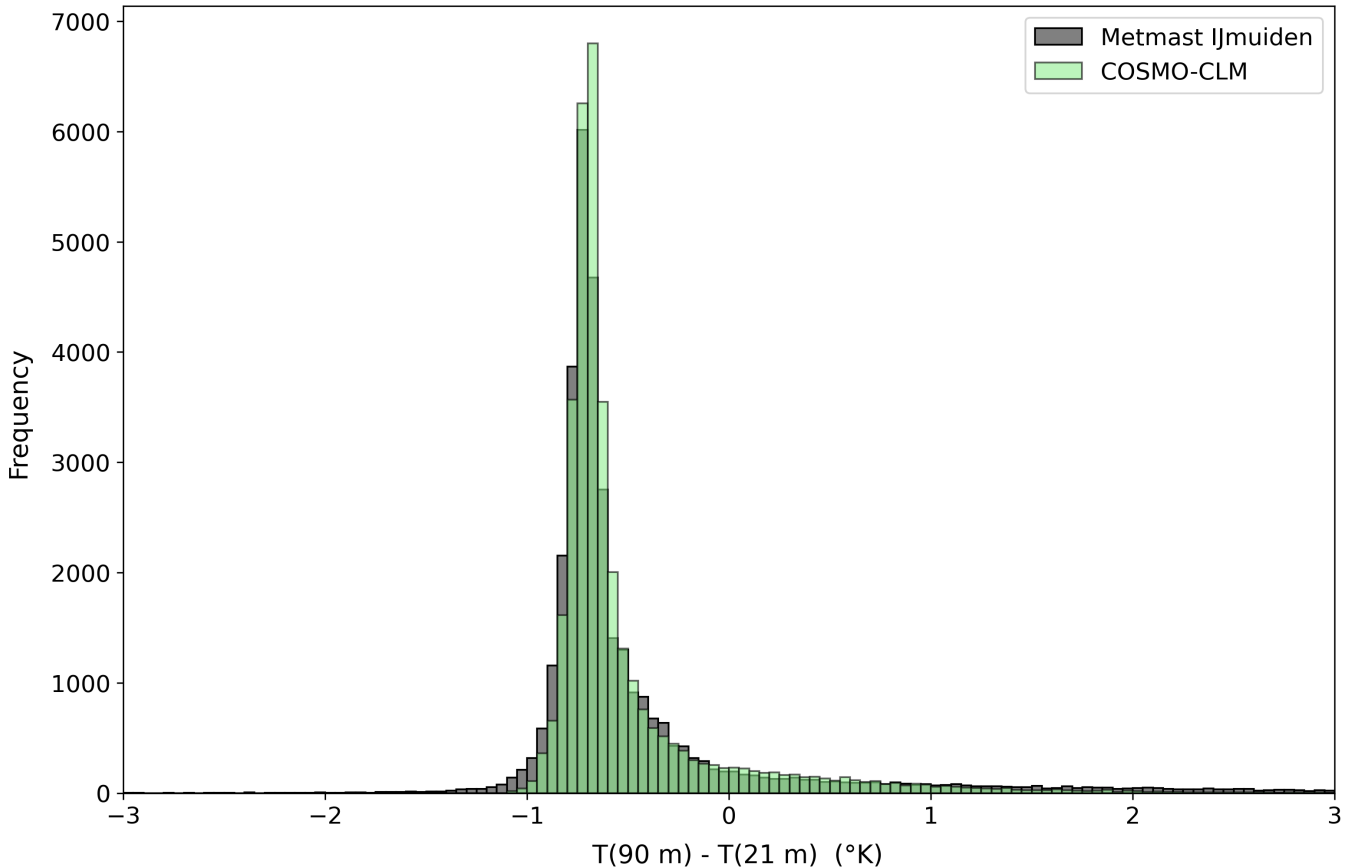


Figure 1: Histograms of the instantaneous temperature gradient between 90 m MSL and 21 m MSL based on 10-minute data for the period 2012-2015 at the location of measurement mast Ijmuiden (MMIJ). Lightgreen: only COSMO-CLM; Grey: only the measurements; Green: overlap between the histograms.

In addition to this, another sentence is added right after this which mentions the good temporal correlation of modelled and measured temperature gradients (Pearson correlation coefficient of 0.85) to warrant our choice to use the same timesteps for each stability class in the measurements as were determined based on the stability computations for the model:

*Because vertical profiles of pressure and temperature are generally not available over the range of the meteorological masts or wind lidar scanning ranges, the stability criterion can only be computed for the model. Based on a good temporal correlation between the temperature gradients of COSMO-CLM and measurement mast MMIJ (Pearson correlation coefficient = 0.85), the timesteps matched to a stability class for the model grid cell nearest to each measurement location were also matched to that stability class for the measurement data.*

**Comment 2:** I am missing a discussion on validating the wake farm parameterisation used. It is challenging to validate these parameterisations in terms of far wakes due to the lack of wind farm data and the fact that current wind farms are not yet as large as the ones you simulate. Please indicate an uncertainty based on the literature. Volker et al. (2015), Fischereit et al. (2022) and other publications show considerable differences between Fitch and EWP schemes with limited validation data, which are relatively close to the wind farm. How much would this uncertainty affect your CF reductions for the North Sea?

**Update:** This very important aspect deserved more attention in our text. Based on an assessment of literature, we find indeed a significant variation across WFP schemes, but also that the evaluations indicate a very good performance of Fitch in terms of modelled wind speed deficits in and behind wind farms, often outperforming other WFP schemes. Two sections were added to the text body which discusses this based on literature. First, in the methods section:

*Several other wind farm parametrizations exist (Fischereit et al., 2022) and it has been shown that the modelled wind speed deficits inside and behind a wind farm can vary substantially from the Fitch WFP (Ali et al., 2023). However, validation of the Fitch WFP with offshore masts, lidars and airborne measurements in the wake of a wind farm has shown very good performance for HARMONIE-AROME as wind speed biases are strongly reduced (Van Stratum et al., 2022; Dirksen et al., 2022). This good performance has also been determined in WRF by comparing to offshore masts*

(Garcia-Santiago et al., 2022) and in COSMO-CLM by comparing to LES (Chatterjee et al., 2016) and airborne measurements (Akhtar et al., 2021). Also wind speed reductions inside of a wind farm have been shown to agree well with airborne measurements (Ali et al., 2023), mast measurements (Dirksen et al., 2022) and RANS simulations (Fischereit et al., 2021). Moreover, comparisons with other WFP schemes show that Fitch generally outperforms these other schemes, both inside a wind farm and in the farm wake (Fischereit et al., 2021; Ali et al., 2023).

And also in the conclusion section, this point is adressed again with the following part:

Whereas comparisons between wind farm parametrizations have shown large variations in terms of modelled wind speed deficits inside and behind wind farms (Ali et al., 2023), validation efforts in several mesoscale models have indicated a very good performance of the Fitch WFP (Fischereit et al., 2021; Van Stratum et al., 2022; Ali et al., 2023). Nonetheless, the use of other WFP schemes might significantly alter the magnitudes presented here, more so due to the large clusters and large wind farms included in the considered layout which can even lead to wake losses for background wind speeds well above rated. Hence, further benchmarking studies of WFP's for a range of atmospheric conditions and validation data could help in further reducing this WFP-related uncertainty. An additional complication here is that this study includes wind farms of non-existent sizes for which validations simply do not exist.

## Minor comments

**Comment 1:** Please follow the WES guidelines for units (e.g., m/s is not acceptable)

**Update:** All the units have now been adapted according to the guidelines.

**Comment 2:** L22: change are to is

**Update:** After inspection, "are" still seems like the adequate option here since there are two subjects, so it was retained.

**Comment 3:** L28: "and gigawatt-scale wind farms emerge..." I find that maybe the verb tense is not right, future?

**Update:** This sentence has been adapted to be more clear in general, thereby also improving the tense use. The updated version is: "Currently, limited space and the urgent decarbonization of electricity systems lead to the installation and planning of very dense wind farms (capacity density  $> 10 \text{ MW km}^{-2}$ ) and exceptionally large wind farms (capacity  $> 1 \text{ GW}$ ) that are strongly impacted by these turbine interactions (citations)."

**Comment 4:** L94. I would add, "However, wind farms have increasingly affected some of the masts used in the validation. " This is the case for FINO1 and FINO3.

**Update:** This is indeed only mentioned shortly in this part in the text. However, in section 2.4.1 we go into more detail about filtering of observations for wind farm disturbance and refer to the corresponding supplementary table which summarizes filtered angles or periods per station. So, we decided not to add another mention of this here.

**Comment 5:** Some of your symbols are sometimes italics and sometimes not. e.g. R in 104. Also, after the equations. All symbols should be in italics.

**Update:** This has now been streamlined across the complete text: all variables are now consistently italicized in the text body.

**Comment 6:** Is there a direct relationship between the PSS and the EMD? I have a feeling it does. It could be good to mention and thus be able to compare your statistics with those of the NEWA simulations.

**Update:** The relationship between the PSS and the EMD metric has now been added to the text to support comparability between these metrics. The following mention was added to line 113 of the initial submission: "... For one-dimensional histograms, this metric is connected to the Earth Mover's Distance (EMD) metric, which in contrast represents the area of mismatch between two histograms (Rabin et al., 2008)"

**Comment 7:** L211: This is not an extrapolation, right? The values above and below the sensor height are known. BTW, this is analogous to a log interpolation of the wind speed between the levels. Too much text confuses people. Also, we want to move away from using power law relationships when data for interpolation is available. The way you write this method could give the wrong impression.

**Update:** The word "extrapolation" was indeed updated to "interpolation". As discussed in the reply to the referee, a test has indicated negligible differences between both interpolation methods so that there are no implications for the findings of our study. Adding a discussion on this topic in the text body seems out of the scope of this manuscript, definitely since the reply to the referee on this topic is publicly available.

**Comment 8:** L241. Please use conventional abbreviations (for example, from textbooks) for often-used quantities, e.g.  $Ri_B$ .

**Update:** Adapted for the Bulk Richardson number throughout the text based on the textbook of Stull (1988) to  $R_B$ .

**Comment 9:** Captions of Fig 8-10: Grey shadings represent wind farm locations.

**Update:** This has been updated for all relevant captions

**Comment 10:** How are the turbines located in each grid cell? That should be mentioned.

**Update:** To include this information, a sentence was added after line 105 of the initial submission: "Because the wind farm parametrization assumes that turbines within a single grid cell never have any wake interactions, no additional information is required on the layout of the turbines in each wind farm."

**Comment 11:** Extrapolation of each measurement has been shown to work poorly. See Badger et al. (2016) in JAMC, DOI:10.1175/JAMC-D-15-0197.1

**Update:** It is true that this constant coefficient extrapolation is too simplistic as pointed out by the referenced publication, because stratification effects are not taken into account. This is a limitation of this part of the evaluation component in our study, but we have tried to mitigate this by selecting the stations with a measurement height below 30 meters height, so close to the target height of 10 m, in order to reduce the impact of extrapolation errors. While there are stations in our study with even lower measurement heights (i.e. < 20 m), these are located too close to the coast, so that no ASCAT data is available in these locations.

**Comment 12:** L258-259: Please be more explicit on what data is used to compute the static stability

**Update:** Additional details have been provided on the amount of model levels in the considered height range. Also more details were provided on the stability subdivision of the measurement data.

**Comment 13:** Figure 2. Please explicitly name the period used.

**Update:** The description in the caption was adapted to make more clear that the data is for the full measurement period of each individual station and a reference to the tables holding this period information was included in the caption. Adding the individual time periods to the figure would also make the figure more messy, we think.

**Comment 14:** Figure 7. The height at which the maps are computed and the time period details are also missing.

**Update:** The heights have been explicitly stated now for all relevant figures (also supplementary).

**Comment 15:** L360-361. It is not clear what you mean by “data points”. Are these in space or in time? I guess in time, but maybe a time match will be best or a frequency.

**Update:** This concerned time points. For more clarity, frequencies were now adopted as suggested by the referee.

**Comment 16:** I like the transects in Fig. 8. Especially because they help assess the necessary spacing between wind farm clusters regarding recovery distances. But again, you are missing the height in Fig 8 caption. Please make sure that the captions are complete.

**Update:** For figures 8 and 9 and Table 4 these heights have now been added explicitly in the captions.