Reply to CC1:

Dear authors,

thanks for the very interesting manuscript. I have two comments:

Q1. Beside the references you mentioned for a model description of COSMO6.0, the usual (or default) references additionally include:

- sub-grid scale orography (SSO) (Lott and Miller, 1997; Schulz, 2008)

- TERRA (Schrodin and Heise, 2001; Schulz et al., 2016; Schulz and Vogel, 2020)

Maybe you could also say a word whether the updated parameterizations were only available in the latest model version COSMO6.0, or they were really switched on and used. The model performance depends very much on this.

References:

Schulz, J.-P., 2008: Introducing sub-grid scale orographic effects in the COSMO model, COSMO Newsletter, 9, 29–36. (Available at http://www.cosmo-model.org/)

Schulz, J.-P., G. Vogel, C. Becker, S. Kothe, U. Rummel and B. Ahrens, 2016: Evaluation of the ground heat flux simulated by a multi-layer land surface scheme using high-quality observations at grass land and bare soil, Meteor. Z., 25, 607–620.

Schulz, J.-P. and G. Vogel, 2020: Improving the processes in the land surface scheme TERRA: Bare soil evaporation and skin temperature, Atmosphere, 11, 513.

Reply to CC1-Q1:

Thank you for your comments. We agree that the TERRA scheme and the land–atmosphere interaction processes are fundamental components of the COSMO model and represent an essential part of its general description. In our original approach, we chose to highlight only those model schemes most relevant to our experiments, in order to maintain focus and clarity. This decision was motivated by the fact that our atmospheric domain primarily targets the North Sea region, where roughly 60% of the area is covered by water, and our analysis is limited to model performance over open-water areas. Nevertheless, we acknowledge that a more comprehensive model overview should also include a reference to the TERRA scheme and the corresponding literature.

In section 3.1, Schulz (2008) will be added to the SSO literature and TERRA scheme will be included as follows:

The land–atmosphere interaction is represented by the TERRA multi-layer soil–vegetation–atmosphere transfer scheme, which explicitly resolves heat and moisture transport within soil, snow, and canopy layers. TERRA ensures closure of the surface energy and water balance by simulating soil temperature and moisture dynamics, evapotranspiration, snowpack evolution, and the surface fluxes of heat, moisture, and momentum. (chrodin and Heise, 2001; Schulz et al., 2016; Schulz and Vogel, 2020).

Regarding the updated parametrization: The reimplementation of Fitch et al. (2012) in COSMO 6.0 was developed specifically for this research and therefore cannot be included in the standard COSMO distribution, since COSMO 6.0 was the last official release. Instead, the reimplementation is made openly available as a separate module, published as a patch for COSMO 6.0 in the Zenodo repository (Elizalde, 2023). After applying the patch, the parametrization can be activated through a dedicated switch within the standard model configuration settings.

In section 3.2, the first paragraph will also include the description:

The reimplementation, along with the new updates, is provided as a separate module, publicly available as a patch for the COSMO6.0-clm code in the Zenodo repository (Elizalde, 2023), and it can be activated through a new dedicated switch within the standard model configuration settings.

In ‘Code availability’ Section, the text will be modified:

The reimplementation of the wind farm parameterization of Fitch et al. (2012), along with the new updates, is provided as a separate module, publicly available as a patch for COSMO 6.0 in the Zenodo repository (Elizalde, 2023).

Reference:
Elizalde, A. (2023). Wind farm parametrization for COSMO6.0-clm. Zenodo. https://doi.org/10.5281/zenodo.10069391

Fitch, A. C., Olson, J. B., Lundquist, J. K., Dudhia, J., Gupta, A. K., Michalakes, J., and Barstad, I.: Local and Mesoscale Impacts of Wind Farms as Parameterized in a Mesoscale NWP Model, Monthly Weather Review, 140, 3017–3038, https://doi.org/10.1175/mwr-d-11-00352.1, 2012

Q2. Maybe you could discuss a bit if you want to transfer this work and continue it to COSMO's successor model ICON.

Reply to CC1-Q2:

To the best of our knowledge, DWD will proceed with the implementation of Fitch et al. (2012) scheme in ICON.

Reply to RC1

Review of “Uncertainty in Offshore Wind Power Forecasts: A Regional Climate Modeling Approach for the North Sea” by Alberto Elizalde, Naveed Akhtar, Beate Geyer, and Corinna Schrum, submitted to Wind Energy Science

The manuscript describes the modeling efforts by the team to quantify the uncertainty in future wind power production as predicted with the COSMO-CLM model over the North Sea, focusing on uncertainties due to the choice of the reanalysis dataset for initial and boundary conditions (6%), the turbine density (2%), and the wind speed at different turbine heights (5%). The uncertainties as defined in the paper are hard to fully understand and somewhat questionable in their interpretations, as discussed in more detail below.

The manuscript is exceptionally well written and logically structured, a pleasure to read really; the scientific approach is solid and the results are relevant, although it is difficult to tell whether they are extendable to other forecasting tools or they are specific to the COSMO-CLM alone. An important innovation of this study is the use of different turbine models and types for the various wind farms, as close as possible to the actual installed type for existing farms or to the most advanced type for the planned ones. As far as I know, this has not been done before and it is a valid approach.

In my opinion, the paper should be published after the issues below are addressed.

Reply: We sincerely thank the reviewer for their detailed and constructive comments, which have helped us improve the manuscript. We now address their remarks point by point below:

Major issues

1. L. 160: there are intense discussions in the community about the correction factor for C_TKE. Ignoring it entirely seems untimely. Even though its effects on power production are probably small (based on conference talks I attended), the uncertainty introduced might be of the order of 2-5%, similar to the magnitude of the uncertainties addressed here. I recommend running at least a period per season (with just one reanalysis perhaps) with a reduced value of C_TKE to have a sense of its impact.

Reply: We agree with the point raised by the reviewer. Our original intention was not to introduce further modifications to the parametrization itself, but rather to extend its application to power calculations and to demonstrate the model’s new capability of handling non-homogeneous wind farm configurations. However, as the reviewer correctly points out, recent studies have been found that the value of C_{TKE} represents a source of uncertainty in Fitch et al. (2012) parametrization, and its influence on power output is indeed an important and timely topic. To address this, we performed an additional two-year simulation with C_{TKE} reduced to one quarter of its original value, as proposed by Archer et al. (2020). This simulation was then compared with its equivalent setup using the original C_{TKE} value over the same period. We added a new Subsection 4.6 to the manuscript that analyses the effects on wind speed and explains the associated reduction in power output.

2. Table 4 and relevant text: the results of the Chronological case are dramatically different (and better according to observations) than those from the other runs in 2012-2018. I thought that this did not make sense, then I realized that the authors did not conduct a correct validation here and therefore should redo this part. Two wind farms were installed in 2015 near FINO1 and the Chronological run had no wind farms from 2012 to 2015, thus low bias, whereas the other runs incorrectly had those wind farms over those 4 years, thus high bias. Similarly at FINO3, 3 wind farms were installed in 2015 and 2017, thus similar issues. The validation must be conducted only during the last year, 2018, in which no new wind farms were added. My expectation is that the biases will be very similar among the various runs.

Reply: We understand the reviewer’s concerns. We are fully aware of the temporal mismatch between wind farm development around the FINO platforms and the setup of most of our simulations (with the exception of the Chronological run). This discrepancy is intentional, as our simulations are designed as theoretical, technical-scenario-based experiments. All technical scenarios prescribe the entire set of wind farms included in the EMODnet dataset from the start of the simulation, regardless of their actual commissioning year. This means that, in addition to operational farms, areas under construction, planned, or already approved are also included. This enables us to assess the dependency of the impact of the wind farms on the atmosphere and the power out more or less undepend on the weather situation. It is necessary due to the high year to year variability of the wind climate. We anticipated that such a setup would lead to an underestimation of wind speed averages compared to the observed at the FINO1 platform, since the simulations incorporate more wind farms than those present in reality at the corresponding time, as noted in line 205:

“… for the scenario simulations—except for the Chronological case—all wind farms are considered operational throughout the entire simulated period. This includes all wind farms in the vicinity of the FINO1 cluster as well as those in the adjacent eastern cluster, which amplifies the effects of wake effects and leads to an additional wind speed reduction …”

It is not the purpose of these simulations to reproduce the wind speed averages observed at FINO1. Rather, our primary objective in using FINO data for statistical comparison is twofold: (i) to illustrate the impact of different wind farm development scenarios on wind speed at a fixed location, and (ii) to quantify the influence of boundary conditions on the simulated wind field. The FINO datasets are well suited for this purpose, as they provide a multi-year time window against which the temporal variability of our simulations can be evaluated. Restricting the comparison to a single year, such as 2018, would not alter the overall conclusions: on the one hand, average wind speeds in the scenarios would still be lower due to the inclusion of more activated wind farm areas than present at the platform; on the other hand, such restriction would reduce the robustness of the statistics regarding temporal variability. For these reasons, we believe the current analysis is appropriate and no further modifications are necessary. Nevertheless, to improve clarity to this matter, we have revised the text at line 205 to read:

“Conversely, in the scenario simulations—except for the Chronological case—all wind farms from the EMODnet dataset are activated at the beginning of the simulations. These setups include not only operational farms but also those under construction, planned, or approved. In this way, all those wind farm areas in the vicinity of the FINO1 cluster as well as the adjacent eastern cluster are included, which amplifies wake interactions and leads to a further reduction in wind speed than observed, …”

3. Table 6 and relevant text: unfortunately the analysis and interpretation of these results is not correct. The difference between the two No-wake runs is not the effect of “wind conditions by taller turbines” (150 m vs. 90 m) alone, because two different rotor diameters D were used too (240 m vs. 120 m). To assess the impact of hub height H alone, the 3.6 MW run should have been compared with one identical but with a hub height of 150 m instead of 90 m. Only that way the only power output difference would have been due to the different wind conditions at higher hub height. The difference value of 5.3% is a combination of hub height and diameter combined, in the absence of wake effects … not sure what that really is, actually. Similarly, the difference between wake losses of -15.5% and -13.5%, thus 2%, is not due to turbine density alone (which would depend on D only), but also on H. All that we can conclude from Table 6 is that the two runs with wakes and different H and D give a higher output by 7.4% for the larger turbine, resulting from a combination of H and D effects that cannot be detangled with the current runs.

Reply: We acknowledge this point, noting that the impact depends not only on hub height but also on rotor diameter. Our intention was to assess the full turbine (hub height and rotor diameter), not hub height in isolation. For brevity, turbines with a 150 m hub height and 240 m rotor diameter were referred to as “taller” turbines, while those with a 90 m hub height and 120 m rotor diameter were referred to as “smaller” turbines; however, only hub height was mentioned, which may have caused confusion. We do not intend to perform a separate analysis on hub height alone, as this would provide limited additional insight. Instead, we revised the manuscript to eliminate any implication that only hub height was considered, clarifying that the full turbine—including hub height and rotor diameter—was accounted for.

Minor issues

4. L. 75-85: about the issue of the power coefficients. Fig. 1 shows the issue well: the power curve (solid line) gives a different power from that calculated by multiplying the hub-height wind speed by the power coefficient at that speed (dash-dotted line) and the corrected curve is the dashed one. There seems to be only one correction factor for the entire curve from Table 1. Wouldn’t it make more sense to correct each value of the power coefficient individually at each wind speed, rather than using one blanket value for all speeds? This is not a big deal and I am not requesting this work, just curiosity and the fact that the weird wiggles and the bumps at high speeds remain and are not too realistic.

Reply: Although computational methods are available to calculate a turbine’s power and thrust coefficients, manufacturers typically obtain these values through a combination of numerical modeling and experimental approaches. As a result, there is no straightforward way to determine them for individual wind speeds, since they do not follow a simple formulation. Therefore, our corrections are applied across all values with the aim of ensuring that the turbine’s rated power aligns with its theoretical rated value.

5. L. 99: are ASCAT data once a day only?

Reply: ASCAT provides two measurements per day: one during the ascending pass and one during the descending pass. To fairly compare ASCAT with our model output, we selected only the hours closest to the satellite overpass times, and only these data were used for comparison. For clarity, the sentence in question at line 99 has been rewritten as follows:

“It provides two daily measurements: one from the ascending pass and one from the descending pass.”

6. L. 142: what is the rotated North Pole?

Reply: The model grid is defined using rotated geographical coordinates (line 123), intended to keep the grid boxes as uniform in size as possible over the region of interest, in our case, the North Sea. In this rotated system, the standard geographical North Pole (0° W, 90° N) is shifted to (180° W, 30° N).

7. Table 2 and text: how did you calculate the power without wake losses in the "Control 3.6 MW" and "Control 15 MW" scenarios? I suspect it is just the Control run post-processed to calculate an (unrealistic) output offline if all turbines were front-row.

Reply: No post-processing was applied, as the model output wind speed as hourly averages or as instantaneous values. To calculate the equivalent power output to the other simulation but without wake losses, the model was rerun with the full wind farm parameterization activated, but within the model code, the lines that compute TKE and momentum terms were commented out, thereby neglecting their influence. As in the other runs, the power output was directly computed at each model time step and stored as hourly accumulated values. While the difference between rerunning the model in this manner and estimating power output in a post-processing step may not be substantial, we consider rerunning the model to be the more robust and consistent approach. For clarification, line 337 has been revised to read:

"To isolate the individual impacts of turbine density and turbine sizes, we conducted two additional ERA5-driven simulations for the 3.6MW and 15MW scenarios. In these simulations, generated power was calculated by the model under unperturbed flow, achieved by neglecting the contributions from the momentum sink and TKE source terms."

8. L. 164-167: please explain how the large (15 MW) and small (3.6 MW) turbines were arranged over the same wind farm areas for the 3.6 MW and 15 MW cases. What spacing (8Dx8D? 10Dx10D?) was assumed?

Reply: Since the model resolution in our case is 2 km x 2 km, smaller-scale processes cannot be explicitly resolved. The parametrization of Fitch et al. (2012) assumes that all wind turbines within a grid box are effectively located at its center in superposition, without interacting with one another. In other words, if the turbine density for a specific wind farm (N_T in our formulation) is sufficiently high that multiple turbines fall within a single grid box, their mutual wake effects are neglected. For clarification, line 157 has been revised to read:

“N_T is the turbine density, calculated as the number of turbines per wind farm area, assumed to be located at the center of the grid box in superposition, with no mutual interaction, …”

9. L. 175 and other occurrences throughout: using two different (but not entirely independent) datasets like ERA5 and ERA Interim does not quantify the uncertainty due to “atmospheric conditions”, rather that due to the IC/BC. Basically, any bias in the model chosen for IC/BC is transferred to the simulations, thus this uncertainty reflects the importance of the IC/BC. Please correct the text throughout including the abstract.

Reply: We agree with this observation. The manuscript has been corrected to clarify in the relevant text that we are referring to boundary conditions, rather than atmospheric conditions, in this context.

10. L. 187: show the equation for the bulk Richardson number and explain how you obtained all its terms from the model values and levels.

Reply: The bulk Richardson number is computed internally by the COSMO model at every time step and documented in COSMO model manuals (Doms et al., 2021). No post-processing was applied. The methodology for calculating the bulk Richardson number is thoroughly documented in the COSMO model manuals. However, we have included the equation and an explanation of the terms in Subsection 3.4 as suggested.

11. Table 4: this table must be redone as mentioned at item 2) above, but if the p-value equals 0.0 for all cases, then it does not need to be shown. However, I recommend using more digits, at least 2 after the decimal separator.

Reply: In Table 4, the p-values have been removed, and a comment on the caption has been added indicating that “All correlations were found to be statistically significant with p-values below 0.01.”

12. Figure 5 and relevant text: a definition of what the “anomalies for wind speed” are is necessary. Without it, it appears that the results of the simulation were rather poor and that the wake was exaggerated in lateral extent and length. Please describe what you are showing (anomalies with respect to what?) and why this case was chosen, keeping in mind that it is not convincing of the performance of the wind farm parameterization.

Reply: In climate data analysis, anomalies refer to deviations of a dataset from a reference value, typically its long-term average. For clarity, Figure 5 caption and related text have been revised to read: “Wind speed anomalies, calculated as deviations from the respective dataset mean ...”

13. Figure 7b: this flow does not appear to be from the 183-193 direction. In addition, these results appear very noisy, with huge (unrealistic) accelerations to the north east of the big wind farm cluster. Perhaps numerical noise?

Reply: The direction range of 183° to 193° does not represent the wind field direction of each panel in the figure. As stated in line 284 “*Figure 7 shows composites of wind speed deficits based on different atmospheric conditions*”. *Each* composite is formed from the average of cases (constituents) filtered according to wind direction and atmospheric stability criteria applied only at the Amrumbank West wind farm—the northernmost farm of the southern cluster within the delimited area marked by the red line—not from the entire wind field. The wind direction criterion, described in line 286, requires that “*the wind direction should have a southerly component within ±5° centered at δ = 188°, i.e., in the range 183°–193°.*” Outside the Amrumbank West wind farm, the wind direction can be any.

Since each composite is constructed from the average of constituents at different time steps and under different atmospheric conditions, the resulting composite does not necessarily depict a physically realistic wind field. However, the averaged pattern of the composite provides an indication of the typical tendencies of the atmospheric state under the selected criteria. Line 286 has been amended for clarity as follows:

“*The wind direction at Amrumbank West wind farm should have a southerly component within ±5° centered at δ = 188°, i.e., in the range 183°–193°.*”

14. L. 283 and a few other instances: there is no way an atmospheric flow could be laminar, the Reynolds’ numbers are of the order of millions.

Reply: All instances where the term “laminar flow” was used have been reviewed and revised accordingly.

15. L. 370-371: the wake results were definitely not “well represented”, see also comment 12) above.

Reply: While the wake extent would not be captured with the accuracy of LES models, for a regional mesoscale model such as COSMO—which operates at coarser resolution and parameterizes sub-grid processes—the wake extent is relatively well represented, providing a realistic depiction of mesoscale wake effects. Line 370 is now amended to state:

"*The model reasonably captures the vertical structure and extent of the wake produced by the wind farms, considering the resolution and scale of the mesoscale model.*"

16. L. 383: should 58 GW be 62.6 GW?

Reply: The calculated annual power output of 58.9 GW corresponds to the 3.6 MW scenario driven with ERA-Interim forcings. This result was mistakenly omitted but is now displayed in Table 6. Based on this, the power output uncertainty derived from our set of simulations amounts to approximately 15 GW, spanning a range from about 59 GW (3.6 MW ERA-Interim simulation) to about 74 GW (15 MW ERA-5 driven simulation).

17. L. 380-383: we already discussed that the actual values of the uncertainties here are not correct and do not really represent what is stated. Nonetheless, there seems to be an assumption here that the uncertainties can be linearly summed up, 5+2+6=13%, which I am not sure about. In fact the power output difference is 74-63=11 GW, which I am not sure about. In fact which is 11/150=7.3%, not 13%. Even if we kept the wrong value of 58 MW, then 74-58=16 GW, not 20 GW as stated on l. 383. All these inconsistencies are due to the wrong interpretations and calculation of the individual uncertainties and to the wrong assumption that they would be linear.

Reply: Subsection 4.5 and the related text throughout the manuscript have been revised to provide a correct explanation, and the calculation of percentages attributable to driving conditions and turbine types is now presented as follows:

Our analysis indicates that the combined uncertainty due to driving conditions and the difference of turbine types amounts to approximately 15 GW, corresponding to 10% of the total installed capacity (150 GW). Of this, the contribution from driving conditions accounts for about 4 GW (2.5% w.r.t. 150 GW), whereas the difference of turbine types has a larger influence, contributing roughly 11 GW (7.5%). Further breaking down this last contribution by separating the effects of turbine size and wake interactions, we find that, when neglecting wake effects, the difference in turbine size accounts for approximately 8 GW (5.3%), while the remaining 3 GW (2.2%) is attributed to difference of wake effects depending on the turbine size.

The TKE production sensitivity experiments indicate that reducing C_{TKE} to one quarter of the original value decreases TKE production at hub height within wind farm areas by approximately 0.5 $m^2 s^{-2}$. This reduction translates into a decrease in wind speed of about 0.2 $m s^{-1}$ a corresponding reduction in generated power of roughly 2 GW (1.4% w.r.t. 150 GW), equivalent to a 1.4% decrease in the overall capacity factor.

Refences:

Archer, C. L., Wu, S., Ma, Y., and Jiménez, P. A.: Two Corrections for Turbulent Kinetic Energy Generated by Wind Farms in the WRF Model, Monthly Weather Review, 148, 4823–4835, https://doi.org/10.1175/mwr-d-20-0097.1, 2020.

Doms, G., Förstner, J., Heise, E., Herzog, H.-J., Mironov, D., Raschendorfer, M., Reinhardt, T., Ritter, B., Schrodin, R., Schulz, J.-P., and Vo- gel, G.: COSMO-Model Version 6.00: A Description of the Nonhydrostatic Regional COSMO-Model - Part II: Physical Parametrizations, Tech. rep., Deutscher Wetterdienst, https://doi.org/10.5676/DWD_PUB/NWV/COSMO-DOC_6.00_II, 2021

Fitch, A. C., Olson, J. B., Lundquist, J. K., Dudhia, J., Gupta, A. K., Michalakes, J., and Barstad, I.: Local and Mesoscale Impacts of Wind Farms as Parameterized in a Mesoscale NWP Model, Monthly Weather Review, 140, 3017–3038, https://doi.org/10.1175/mwr-d-11-00352.1, 2012.

Reply to RC2:

This is an interesting study on regional climate modelling of wind power in the North Sea. I do however have an issue with the paper, in particularly with the claim that it provides an uncertainty assessment of offshore wind farm power forecasts, which, in my opinion, it does not deliver. In this paper, the uncertainty assessment essentially boils down to evaluating the impact of different driving data for the regional climate model—namely ERA5 and ERA-Interim—and secondly, to comparing the implementation of 3.6 MW and 15 MW turbines at the same capacity density.

In my view, this does not constitute an uncertainty assessment. Firstly, I do not consider the sensitivity of different turbine types at the same capacity density to be part of the uncertainty. A wind farm developer typically knows which type of turbine they intend to use. And even if there is uncertainty, then the 3.6 MW turbine is much to small to consider for future parks.

We sincerely thank the reviewer for their detailed and constructive comments, which have helped us improve the manuscript. We now address their remarks point by point below:

It is important to note that for long-term wind farm planning, often spanning decades, the exact turbine types are typically not finalized at the early stages of development. While developers may have a general range of turbine capacities in mind, specific models, rotor diameters, and hub heights are subject to technological advancements, supply chain availability, and economic considerations at the time of construction. Additionally, when new wind farms are installed within existing clusters, the exact performance of both the new and neighbouring farms—in terms of atmospheric circulation, wake interactions, and power generation—cannot be precisely predicted in advance (Finserås et al. 2024). Therefore, contrary to the reviewer's comment, uncertainty in turbine types represents a genuine component of the overall power generation uncertainty and should not be considered solely a sensitivity study.

We are fully aware that the use of a 3.6 MW turbine is not realistic for future offshore wind farms; however, the two rated powers considered in the scenarios (3.6 MW and 15 MW) were intentionally chosen to represent, in purely theoretical scenarios, the extreme ends of the turbine size spectrum. This approach allows us to quantify the maximum range of uncertainty associated with turbine type and layout, rather than to suggest that a 3.6 MW turbine would actually be deployed. By examining these bounds, our analysis captures the potential variability in power generation due to turbine characteristics, providing an estimate of uncertainty that is relevant for long-term planning and assessment of wind farm clusters.

Secondly, the driving data are not the only source of uncertainty. More important sources of uncertainty include: 1) Uncertainty in the background wind fields without wind farms, which can be much broader than the sensitivity to different reanalysis datasets – regional climate models also have biases. 2) The wake effects within and between wind farms, which are subject to uncertainties, even if the turbine type and capacity density is known, the wakes are uncertain. 3) Deviations of actual turbine power curves from the industry-specified ones. And 4) Downtime of the turbines due to curtailment and maintenance.

We acknowledge that additional sources of uncertainty exist and agree that further research on their impacts would be highly valuable. We address the comments point by point below:

1) To our knowledge, few studies have specifically compared wind speeds over the North Sea using different regional climate models (RCMs) driven by the same reanalysis boundary conditions, as relevant to the reviewer's comment. Within the Euro-CORDEX framework, Moemken et al. (2018) compared the RCA and COSMO RCMs, both forced with ERA-Interim. Their results (their Figure 1) indicate that the North Sea wind speed fields are very similar, with differences of less than 1 m s^{-1}. In contrast, background wind fields from RCMs driven by different boundary conditions derived from future climate projections exhibit substantially larger uncertainty, primarily due to differences in the driving conditions. Ganske et al. (2016) reported that the 50th percentile range of wind speed among RCMs over the period 1961–2099 is approximately 4 m s^{-1}. While it is feasible to estimate power output directly from existing wind datasets, this is not the focus of the present study. Instead, our objective is to investigate the dynamical interactions of wake effects within wind farm clusters by explicitly calculating wakes at each model timestep, thereby capturing their continuous evolution throughout the simulation and assessing the resulting impacts on power generation.

2) Part of our investigation is precisely to investigate wake effects within and between wind farms as stated in our abstract (line 6): "*However, the uncertainty of such assessments has not been fully addressed. Wake effects have been identified as the primary source of power losses. They are often studied within individual wind farms or small clusters, but the dynamics of large wind farm clusters at a regional scale are only beginning to be explored. In this study, we address uncertainties of power output derived from projected wind farm areas at the North Sea in scenarios that encompass different turbine setups and driving conditions.*". Specifically, Subsection 4.5 (renamed 'Uncertainty in Wind Farm Power Production') details our findings on this topic.

3) & 4) We acknowledge that the lack of real turbine power and thrust coefficients, as well as turbine downtime and maintenance, contribute to total power uncertainty. However, these data are typically proprietary and not publicly available due to company economical strategies, making it currently infeasible to include them in our assessment.

Lastly, it is not at all clear whether ERA5 and ERA-Interim represent the full range of uncertainty found in state-of-the-art reanalysis products since there are others such as COSMO-REA, NCEP and MERRA. The winds in other re-analyses products over the region of interest should be discussed at the minimum.

We agree that ERA5 and ERA-Interim do not represent the full range of uncertainty for model driving conditions. We included a discussion to this respect at line 56 to read:

"With respect to boundary conditions, Hahmann et al. (2022) compared wind speed estimates from four reanalysis products (NEWA, ERA5, MERRA2, and 20CRv3) with mast observations from several offshore platforms in the North Sea (FINO1, FINO2, FINO3, IJmuiden, and Ekofisk) over the period from 1980 to 2014. Their analysis showed that, when evaluated in relation to the observational data (see their Fig. 2), the wind climates produced by the different reanalyses were "nearly identical," with mean values ranging from 10.0 to 10.2 m s^{-1}. To investigate for these small uncertainties associated with the choice of boundary conditions and their impact on power production, our scenarios were driven using two different reanalysis datasets (ERA-Interim and ERA5)."

Therefore, I believe the paper needs to be reworked and the scope should be changed. The most important point for improvement is that the paper should not claim to provide a uncertainty assessment, because that is not what it actually does, it is rather a sensitivity analysis to driving re-analysis data and the effect of using different (spatially homogeneous and non homogeneous) turbine types at the same capacity density.

We respectfully disagree with the reviewer's assessment that our study constitutes only a sensitivity analysis rather than an uncertainty assessment. In our view, the sources we investigate indeed represent fundamental uncertainties. Specifically, it is uncertain which turbine types will be installed in the future (a decade or more), how they will be configured, and what capacity densities will be employed. These design choices inherently affect wake dynamics, which themselves constitute an important source of uncertainty. In addition, future wind speed conditions over the North Sea are subject to considerable uncertainty arising primarily from the unpredictable development, tracks, and intensities of low-pressure systems, as well as the complex atmospheric processes influencing them. As a concrete example from our study, the reanalysis datasets employed do not agree on the magnitude of the average wind speed over the North Sea. This average wind speed is therefore itself uncertain, as demonstrated in our comparison with satellite observations, which also may contain errors and uncertanities (see Subsection 4.1, *Wind speed validation*).

We are not examining changes from single input parameters or assumptions (except in our new Subsection 4.6). We are quantifying the range of possible outcomes given unknown or unknowable future conditions. Taken together, these factors demonstrate that our work goes beyond sensitivity testing and addresses genuine uncertainties in future wind power projections.

A second important point for improvement is the presentation of key numbers in the paper:

1) The paper states that increasing the turbine nominal power from 3.6 MW to 15 MW leads to a 5% increase in power production when wakes are not accounted for. In the conclusion, this is presented as part of the uncertainty attributed to turbine type. However, treating this full range as uncertainty implies we have no information about which turbines will be installed—which is not the case. Thus, this is not a true uncertainty range but rather a sensitivity analysis (see above) and should be presented in that way. In the abstract, it is written that wind regimes at different hub heights contribute an additional 2%. Apart from being confusing, I believe the authors may have mixed up the numbers, and it should actually be 5%.

Addressed in the next reply.

2) The paper identifies a 2% stronger reduction in power output due to wake effects when comparing 3.6 MW and 15 MW turbines. This is however not correct – it should be either 2%pt or 14% (the latter being (15.5-13.5) / (0.5 * (15.5 + 13.5))) In the conclusion, this is summarized as approximately 2% being derived from wake intensities caused by the turbine density distribution. This sentence is very unclear, and the earlier formulation would be much better to include in the conclusions. In the abstract, it is written: "Our results show that wake dynamics resulting from different turbine density distributions can account for up to 5% of the variability in generated power." However, it is not different turbine density distributions, but rather the change from 3.6 MW to 15 MW turbines while keeping the capacity density constant. This should be written more clearly. Additionally, I believe the authors have mixed up the numbers — should be 2%pt, not 5%pt.

Addressed in the next reply.

In both the abstract and the conclusion, it is stated that the total uncertainty in power output is approximately 13%. In a wind farm scenario with an installed capacity of 150 GW, this results in a power output range from 58 to 74 GW, corresponding to an uncertainty of 20 GW. I assume the authors mean that 20 GW is 13% of the total installed capacity of 150 GW, and this should be made explicit. It is not 13% of the actual power production.

Addressed in the next reply.

Lastly, for the reasons mentioned above, this number of 13% of the installed capacity does not represents the total uncertainty in power output. Many important factors are not included, and there is already knowledge about which turbines will be installed, so this range of turbine types is not a real uncertainty. I cannot recommend acceptation of the paper when it is presented as a total uncertainty of north sea wind farm power production.

We acknowledge the points previously raised. Some of our calculated percentages, and consequently their interpretations, were unfortunately mixed up. We have revised Subsection 4.5 and the corresponding text throughout the manuscript to provide a corrected explanation. In particular, the calculation of percentages attributable to driving conditions and turbine types has been reinterpreted. A concise summary of the changes is provided below:

Our analysis indicates that the combined uncertainty due to driving conditions and the difference of turbine types amounts to approximately 15 GW, corresponding to 10% of the total installed capacity (150 GW). Of this, the contribution from driving conditions accounts for about 4 GW (2.5% w.r.t. 150 GW), whereas the difference of turbine types has a larger influence, contributing roughly 11 GW (7.5%). Further breaking down this last contribution by separating the effects of turbine size and wake interactions, we find that, when neglecting wake effects, the difference in turbine size accounts for approximately 8 GW (5.3%), while the remaining 3 GW (2.2%) is attributed to difference of wake effects depending on the turbine size.

Minor comments:

Regarding the evaluation method, I am not a strong proponent of using RMSE and R^2 for regional climate model evaluation, assuming that spectral nudging is not used to maintain the timing of events. A small displacement—known as the double

penalty—in the position or timing of a weather event can lead to large RMSE and low R^2, even if the model performs reasonably well. Regional climate models should aim to get the statistics right, rather than the exact wind speeds at the right time and moment.

Indeed, similar to spectral nudging, domain size—particularly smaller domains—can influence how closely the model follows observed values. In our case, we do not use spectral nudging; however, the domain is relatively small for the kind of atmospheric model used, so that synoptic-scale low-pressure systems, imposed through the boundary conditions, still exert a strong influence on the interior of the domain. We agree that using RMSE as a metric is challenging for model evaluation, yet despite this, our results demonstrate good temporal agreement with FINO1 observations (Table 4). In addition to RMSE, we also employed other statistical metrics to further assess model performance.

Figure 6: Please change this to show the wind speeds for a certain wind direction interval to get a much more statistically robust signal. The authors have that information available from the long term simulation.

Figure 6 is intended to complement Figure 5 (line 266). Both figures represent snapshots from the day and time of the airborne measurements (10 September 2016 at 09:00 UTC). The measurements were taken under conditions such that the wind direction transported the wake from the southern cluster toward the northern wind farm. The cross-section shown in Figure 6 (marked in red in Figure 5) follows this wind direction and depicts the simulated wind speeds at the moment of the flight. Using data from other dates and wind directions from an interval would not accurately represent the wind speed during the measurements.

There appears to be a displacement of a weather system, with high and low wind speed deficits occurring close to each other—likely from a single event. Please analyze the individual differences between the model runs for different timestamps to determine whether this displacement is due to model error or a real signal caused by the wind farm. If it is a displacement, it should be removed from the sample; if it is a real signal, the authors should explain it.

The reviewer did not specify which figure they were referring to. Based on the comment, we infer that it concerns Figure 7. The explanation for the apparent displacement is as follows:

Figure 7 shows composites of wind speed deficits based on different atmospheric conditions. Each composite is formed from the average of cases (constituents) filtered according to wind direction and atmospheric stability criteria applied only at the Amrumbank West wind farm—the northernmost farm of the southern cluster within the delimited area marked by the red line—not from the entire wind field. The wind direction criterion, described in line 286, requires that “*the wind direction should have a southerly component within ±5° centered at δ = 188°, i.e., in the range 183°–193°.*” Outside the Amrumbank West wind farm, the wind direction can be any.

Since each composite is constructed from the average of constituents at different time steps and under different atmospheric conditions, the resulting composite does not necessarily depict a physically realistic wind field. However, the averaged pattern of the composite provides an indication of the typical tendencies of the atmospheric state under the selected criteria.

Line 321: It is interesting that turbines with a hub height of 150 metres are exposed to more stable atmospheric conditions than turbines with a 90-metre hub height. Explain why. Moreover, discuss that the rotor area spans a large vertical range, so not only the stability at hub height is relevant.

We acknowledge this point, noting that the impact depends not only on hub height but also on rotor diameter. Our intention was to assess the full turbine (hub height and rotor diameter), not hub height in isolation. For brevity, turbines with a 150 m hub height and 240 m rotor diameter were referred to as “taller” turbines, while those with a 90 m hub height and 120 m rotor diameter were referred to as “smaller” turbines; however, only hub height was mentioned, which may have caused confusion. We revisited the manuscript and revised the text where it might have implied that only hub height was considered. In addition, an explanation on variation of stability as height is gain is added. We have revised the text at line 321 to read:

“Turbines with a hub height of 150 m experience a different wind regime, characterized by more frequent stable and very stable conditions compared to turbines with a 90 m hub height. This is due to reduced surface-driven turbulence and persistent temperature gradients at higher altitudes. Near the ground, friction and solar heating generate turbulence and mixing, but these effects weaken with height.”

Line 335: Another important factor is that newer turbines have different power curves, typically with larger regime II. Please assess and quantify the impact of three factors: (1) the smaller wake effects for the 15 MW turbines, (2) the higher positioning of the 15 MW turbines, and (3) the differences in power curves.

Newer turbines achieve higher efficiency through technical and mechanical improvements. However, their thrust and power coefficients are not publicly available, preventing access to their detailed power curves. Consequently, performing a direct comparison is currently beyond our reach. Regarding the quantification between our simulated 3.6 MW and 15 MW turbines, Table 6 provides the annual mean power output for simulations with these turbine types, both neglecting and including wake effects. The quantification of such effects is as follows (starting from point 2):

2) The difference attributed solely to turbine size, with wake effects neglected, is 93.9 GW − 85.9 GW = 8 GW.

1) The difference due to wakes alone is (73.7 GW − 62.6 GW) − 8 GW = 3.1 GW.

3) Differences arising from turbine-specific power curves are intrinsic to the simulated power of each turbine and are reported in Table 6.

Lines 339–355 and beyond: When comparing results to other studies, add the capacity density, which differs between studies. Furthermore, the discussion refers to percentage points. A reduction in capacity factor from 49% to 42% is a reduction of 7 percentage points, not 7%. If it were expressed as a percentage, it would correspond to a 15% reduction in annual energy production or capacity factor. This distinction should be made clear. The paragraph below should also refer to percentage points, not percentages. Check the entire manuscript.

The manuscript has been revisited and revised to update the relevant text regarding the comparison of percentages as the reviewer suggested, correcting them to percentage points, and to include the appropriate capacity densities in the comparison with results from other authors.

References

Finserås, Eirik / Herrera Anchustegui, Ignacio / Cheynet, Etienne / Gebhardt, Cristian Guillermo / Reuder, Joachim Gone with the wind? Wind farm-induced wakes and regulatory gaps 2024-01Marine Policy , Vol. 159 Elsevier BV p. 105897

Ganske, Anette / Tinz, Birger / Rosenhagen, Gudrun / Heinrich, Hartmut Interannual and Multidecadal Changes of Wind Speed and Directions over the North Sea from Climate Model Results 2016-09 Meteorologische Zeitschrift , Vol. 25, No. 4 Schweizerbart p. 463-478

Hahmann, Andrea N. / García-Santiago, Oscar / Peña, Alfredo Current and future wind energy resources in the North Sea according to CMIP6 2022-12 Wind Energy Science , Vol. 7, No. 6 Copernicus GmbH p. 2373-2391

Moemken, Julia / Reyers, Mark / Feldmann, Hendrik / Pinto, Joaquim G. Future Changes of Wind Speed and Wind Energy Potentials in EURO-CORDEX Ensemble Simulations 2018-06 Journal of Geophysical Research: Atmospheres , Vol. 123, No. 12 American Geophysical Union (AGU) p. 6373-6389