Responses to Reviewer Comment

October 27, 2026

We thank the reviewers for their careful reading of the manuscript and for their constructive comments and suggestions. On behalf of all co-authors, we provide below our point-by-point responses to all referee comments.

Reviewer #1

General comments

> *This seems to be an important study with relevance to wind farm planning. So the reviewer started reading with enthusiasm but found himself stuck after a short while.*

Specific comments

Comment #1.C1

> *The Introduction gives a short abstract of existing research citing a few overview papers. The open caveat is analysed and the desired solution is named: a "U-Net" and a "diffusion-based probabilistic surrogate model". Unfortunately, no definition of a U-Net or a diffusion model is given. As these terms seem to define the core of the solution, some sentences on U-Nets and diffusion models should be in order.*
>
> *Due to the missing definition of these two terms the reviewer has not understood the entire second paragraph of Subsection 1.1 (lines 78–90). Section 1 needs reformulation in order to provide access to the content of this manuscript also for non-specialists.*

Authors' Response:

> We agree that brief definitions of U-Net and diffusion models would improve the accessibility of the Introduction. We have revised Section 1, lines 63-65 and 80-82, to provide brief conceptual descriptions of these models, while the detailed model architectures and formulations remain in Section 2. We have also provided background references to the U-Net (lines 150–151) and diffusion models (lines 170-172).

Comment #1.C2

> *The same applies to Section 2 which is totally written in a formal language. Although this might provide a precise notation of the method it is inaccessible for non-specialists. The link to reality seems to be the term "free-stream atmospheric covariates", another term which is not explained (lines 214–220 and Table 2 offer something, but quite late).*

Authors' Response:

The variables such as $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{l}$ are defined at the beginning of Section 2. To further improve readability for non-specialist readers, we have clarified the meaning of "free-stream atmospheric variables" at lines 126-128, and have avoided using the term "covariates" throughout the manuscript. The specific atmospheric input variables are described in detail in Section 3, lines 266-272.

Comment #1.C3

Nothing is said in the introduction whether the authors refer to onshore or offshore wind farms. It is only from the first sentence in Section 3 that the reader learns that the study refers to onshore conditions. It would have been wise to address this aspect much earlier, because wakes are much more intense and larger in offshore environments than in onshore ones. The authors should amend the introduction respectively.

Authors' Response:

The proposed framework is not restricted to onshore wind farms, although the present case study uses an onshore domain. We have clarified this in Section 1, lines 101-103. For offshore applications, the same framework could be used, while the static inputs may differ from those used here. For example, elevation may be omitted or replaced by other relevant static spatial information.

Comment #1.C4

By the way, the method should be applied to the seas around the British Isles or to the German Bight. Here, losses due to wakes are much more relevant than the onshore example given in this study. This would be the true proof of the concept presented.

Authors' Response:

This study was carried out through a collaboration with an industrial partner who provided WRF and WRF-WFP datasets for this onshore domain. We have revised the Acknowledgements section, lines 665-667, to explicitly recognize this contribution.

The framework itself is not restricted to this setting and can definitely be applied to other domains, including offshore regions, when suitable data are available.

Comment #1.C5

Figure 1 should have axes in physical dimensions (kilometres). Otherwise, the given resolution of 1×1 *square kilometre of the WRF model cannot be compared to the size of the wind farms depicted in Figure 1.*

Authors' Response:

We have revised the axes of Figure 1 and Figure B1 to show relative distance in kilometres.

Comment #1.C6

Appendix B can be skipped, as it does not provide any meaningful additional information.

Authors' Response:

Thank you for the suggestion. We have retained Appendix B as it shows the individual wind farm partitioning within each scenario, as this information helps readers

understand the relationship between scenarios and interpret the farm-level results. We have also added introductory text at lines 635-637 to clarify its purpose and relevance in Appendix B.

Comment #1.C7

I stop here making any further detailed comments on this manuscript. The manuscript turns out to be a very technical one and is, unfortunately, not digestible for non-specialists (the reviewer is a meteorologist with special interest in wind energy for decades). Only lines 492–520 in the Conclusions give a little taste how the presented work could refer to the real world of a wind engineer or an energy meteorologist.

Although I think that I understand the importance of the presented research, I can only recommend the publication of this manuscript if the Introduction is extended in order to inform the readers in a more detailed way on the purpose of the study and the linkage between this formal research and the reality. This should start, e.g., by using the term "meteorological variables" instead of "atmospheric covariates". Maybe, a wind engineer or a meteorologist could be added to the team of authors.

Authors' Response:

We recognize the reviewer's concern regarding accessibility. The diffusion-based methods involve several modelling choices and parameterizations that need to be described in sufficient mathematical detail to ensure reproducibility and avoid ambiguity. At the same time, we have revised the manuscript at lines 63-65, 80-82, 123-124, and 126-128 to make it more accessible to meteorologists who may be less familiar with machine learning, while improving the connection between the methodology and its applications in wind energy planning.

Reviewer #2

General comments

This study investigates the use of deep learning models (a U-Net, and two diffusion model variants) to predict regional-scale wind turbine power, accounting for long-range wake effects. The deep learning models are trained using data derived from a mesoscale weather prediction model with a wind farm parameterisation. This addresses a significant problem in wind energy modelling, namely the computational cost of evaluating a large number of layout scenarios using mesoscale models. The study presents a novel approach, and is methodologically sound. I recommend that the paper will be suitable for publication after revisions focusing mostly on methodological clarifications and deeper discussion of the limitations of the current approach.

Specific comments

Comment #2.C1

- *The introduction should make clearer the distinction between farm-scale simulation (where engineering models and most existing deep learning wake modelling approaches are broadly applicable), and regional-scale simulations where mesoscale modelling is appropriate (and as noted by Pryor et al. (2024), is perhaps the only appropriate approach). Here, you also refer to "layouts", which to me implies a farm-scale focus which actually under-sells the benefits of your approach, namely its applicability to long-range interactions.*

Authors' Response:

We agree that the distinction between farm-scale formulation and regional-scale wake modelling should be made more explicit. We have revised lines 28-30 and 33-35 of the Introduction to emphasize that engineering wake models and many data-driven approaches are primarily suited for farm-scale applications, whereas mesoscale modelling is needed to estimate long-range interactions among multiple wind farms over a broader region. We have also clarified that the term "layout" refers to regional-scale arrangements of wind turbines across individual and neighbouring wind farms, rather than only within-farm turbine arrangements. Pryor and Barthelmie (2024) has been included in the Introduction at lines 29-30.

Missing or suggested literature:

Comment #2.C2

- *At lines 32–34, it would be helpful to refer to large-scale studies which have encountered this computational cost problem and which would most benefit from a surrogate model approach; there are numerous examples from Europe e.g. Baas et al. (2024), Dörenkämper et al. (2025), Akhtar et al. (2021).*

Authors' Response:

We have added the suggested references to large-scale wind farm studies that show the computational demands associated with regional atmospheric and wake modelling, and hence the types of applications that could benefit from computationally efficient surrogate models. Please see the revised text at lines 39-41.

Comment #2.C3

- *At lines 75–77, the reasons for contrasting with downscaling problems are not clear. If this is to draw on the literature on deep learning-based downscaling as it relates to the methods you use in this study, this should be expanded.*

Authors' Response:

We have removed the comparison with atmospheric downscaling as it was unnecessary and potentially confusing.

Comment #2.C4

- *I suggest consulting the new review by Dong et al. (2027) for further recent literature on similar problems, in particular refs 98–104 therein, some of which are closely aligned with your study. The novel contributions of your work should be expressed in this context.*

Authors' Response:

We have revised the Introduction to clarify the specific contribution of our study and have included relevant studies cited by Dong et al. (2027). Specifically, we have added the following text in Section 1 (lines 57-70):

> CNN-based models have shown strong performance in wind-related prediction tasks, particularly in multi-site settings (Harrison-Atlas et al.,

2021; Wu et al., 2022; Arslan Tuncar et al., 2024; Liu and Zhang, 2024). More specifically, CNN-based surrogate models are developed for wake prediction at the turbine scale (Wang et al., 2022; Liu et al., 2025a; Santoni et al., 2024) and at wind farm and regional scales (Romero et al., 2024; Weilmann Rasmussen and Réthoré, 2025; Warder et al., 2026). A recent review of machine-learning approaches to wind farm wake modelling is provided by Dong et al. (2027). Among these architectures, U-Net (Ronneberger et al., 2015) is an encoder–decoder convolutional neural network that progressively represents spatial features at different scales and uses skip connections to retain fine-scale spatial information during reconstruction. For example, Santoni et al. (2024) fine-tune a U-Net-based model using large-eddy simulation data to predict high-fidelity velocity fields and capture wake deficits. Using turbine layouts and undisturbed flow conditions, Romero et al. (2024) develop a U-Net to predict wake-affected velocity fields generated by analytical wake models and fluid dynamics simulations. Weilmann Rasmussen and Réthoré (2025) train a U-Net model on numerous turbine layouts simulated using an engineering wake model to predict long-distance wake between wind farms. At the regional scale, Warder et al. (2026) demonstrate the use of U-Net to predict wake deficits given background wind conditions.

In addition, we have also clarified that our study focuses on layout-dependent wind power emulation using paired free-stream and WRF-WFP simulations in lines 87-91, with a systematic evaluation of deterministic surrogate modelling choices and an extension to probabilistic prediction using diffusion models.

Methods and data preparation:

Comment #2.C5

- *Justification should be given for the chosen WRF configuration, e.g. by reference to validation studies. Why is 1 km resolution used, and what is the distribution of the number of turbines per grid cell? This is important to understand the limitations of the power modelling approach.*

Authors' Response:

We thank the reviewer and address the comments in sequence below.

1. We have clarified the rationale for the 1 km WRF resolution in Section 3 lines 217-223. The 1 km grid balances the cost of a full TMY at 10-min output against the accuracy needed for our application, which is assessing the impact of external wakes on a new wind farm. A few kilometres downwind of a farm, as the individual turbine wakes typically have merged into a single farm-scale wake, the resolution needed is set by the farm footprint and inter-farm gaps, not by individual rotors. (Siedersleben et al. (2018), Fischereit et al. (2021), Cañadillas et al. (2022)). The surrogate emulates WRF-WFP as configured, and the revised text says so.
2. We have added the distribution of turbines per grid cell to Section 3 (lines 237-239). In the densest scenario, 555 grid cells contain one turbine, 216 contain two, 57 contain three, and only eight contain four or five turbines. So about 34% of active cells (and about half of all turbines) are in multi-turbine cells.

We agree this matters for the power model, and we had omitted to state that our WRF runs use the multi-turbine axial induction correction of Vollmer et al. (2024). That correction rescales the local grid-cell wind speed to the free-stream value, accounting for the number of turbines in the cell, before computing each turbine's power and thrust. We have added this to Section 3 (lines 226-236) and Table 1.

3. We have expanded the Discussion (lines 593-598) to state that the correction assumes turbines within a cell are identical and experience the free-stream wind, so intra-cell wake losses are not represented. This is expected to slightly overestimate the momentum extracted and power produced in multi-turbine cells (those holding two or three turbines), and only for wind directions aligned with the turbine layout within the cell. Because the far-field wake is governed by the momentum extracted over the whole farm, and most cells hold one or two turbines, we expect the effect on the external wakes of interest to be small. Since the surrogate emulates WRF-WFP as configured, this approximation is inherited by the targets rather than introduced by the surrogate.

Comment #2.C6

- *More information is needed about the types/models of turbines simulated. Are all turbines the same? What is the range of hub heights and rated powers? As much detail as possible should be provided, within the data confidentiality limitations. Turbine capacity range is briefly mentioned in the discussion but should be made clear earlier; it is a key methodological assumption that all turbines can be treated equivalently.*

Authors' Response:

Thanks for pointing this out. The proposed new wind farms considered in this study use the same turbine model under different layouts, while the existing and in-construction farms use a mix of turbine models. Therefore, the case study primarily evaluates generalization across wind farm layouts rather than across turbine types. We have added the ranges of rated power (2.1–4.5 MW), hub height (79–110 m), and total installed capacity (301.5–4480.9 MW) to Section 3 (lines 274-278). We have also clarified that the current capacity grid representation is appropriate for the present setting since turbine characteristics for the proposed farms are held fixed, whereas generalization across different turbine models would require additional turbine-specific information.

Comment #2.C7

- *It should be made clear how the TMY was broken down for the WRF simulations. You mention 3-day simulations in the context of computational cost; is the training also broken down in this way, and how did you arrive at this duration? Is there a spin-up period?*

Authors' Response:

We thank the reviewer and have clarified this in Section 3 (lines 260-265). The TMY is simulated as consecutive 3-day WRF segments, each initialized from ERA5 and preceded by a 6-hour spin-up that is discarded. We use 3-day segments for two reasons. First, in our experience the 1 km inner domain drifts away from the

boundary forcing in longer free-running simulations (beyond about 5 days), so re-initializing keeps the simulations close to the reanalysis. Second, short segments run independently and in parallel, which shortens the wall-clock time for the full TMY. The same segmentation and spin-up are used for the free-stream and WRF-WFP runs, so the paired fields stay consistent. Each month of the TMY is simulated for its representative year as consecutive 3-day segments, with the final segment shortened where needed. The segmentation is purely a simulation device, and the surrogate is trained on the concatenated 10-min series. The models predict each time step independently without using previous time steps as inputs. We have also noted in the Discussion (lines 610-612) that temporal discontinuities between WRF segments would need to be considered in future models incorporating temporal dependencies.

Comment #2.C8

- *The authors should justify, and discuss the implications of, using only a single instantaneous snapshot as input to the surrogate model.*

Authors' Response:

Thanks for raising this point. The surrogate model is formulated to emulate the instantaneous responses of WRF-WFP rather than forecasting future conditions. For each sample, the model takes the contemporaneous free-stream WRF atmospheric fields together with turbine layout and predicts the corresponding WRF-WFP power field at the same time. The atmospheric state provided by WRF serves as the conditioning information for the surrogate, and no previous surrogate outputs are required.

We have clarified this choice and its limitation in the Discussion (lines 577-579 and 607-609). Since no lagged atmospheric inputs are used, the surrogate model does not explicitly account for the history of wake development and advection. This limitation is particularly relevant during rapidly changing wind conditions. For example at a wind speed of 10 m/s, wake disturbances can travel about 6 km in 10 minutes. As a result, the state of a point in a downstream wake may reflect earlier wind directions rather than the instantaneous atmospheric state. In the revised manuscript (lines 604-612), we discuss the potential benefits of incorporating previous atmospheric states and the challenges associated with temporal discontinuities between WRF simulation segments.

The additional wake loss analysis in Appendix D (lines 639-653) also shows that extreme instantaneous errors frequently coincide with rapid wind speed changes, while temporal averaging substantially reduces prediction RMSE. These findings further suggest that incorporating historical atmospheric information may improve predictions under rapidly changing wind conditions.

Comment #2.C9

- *The authors should justify the choice to use ensembles of 10 members for the diffusion models.*

Authors' Response:

We used 10 ensemble members as a practical compromise between computational cost and the objective of demonstrating the probabilistic capability of the diffusion models. The diffusion models are intended as a proof of concept for generating

uncertainty-aware surrogate predictions. Ten members seem sufficient to evaluate ensemble spread, coverage, and CRPS while maintaining manageable computational cost. We also used the fair CRPS to account for the finite ensemble size. We have clarified this choice in lines 316-317 and noted that 10 should not be interpreted as optimal or sufficient to robustly characterize distribution tails. Evaluating the convergence and sensitivity of the uncertainty estimates to larger or smaller ensemble sizes would be useful future work.

Sensitivity analysis:

Comment #2.C10

- *It is not sufficiently clear in Section 3 (and Table 2) what your baseline set of inputs are. E.g. you say "Beyond wind, we consider hub-height air density and turbulent kinetic energy, as well as the potential temperature gradient between 200 m and 20 m." It is not clear that these are excluded from your default model. Lines 354–355 then imply that all variables are in fact included for the "main model comparisons"; please clarify.*

Authors' Response:

Thanks for pointing this out. We have clarified the distinction between the baseline inputs used in the sensitivity analysis and the full input set used in the main model comparisons (lines 266-271). The baseline configuration consists of 100 m wind speed and direction, together with the static layout and elevation inputs. Air density at 100 m, TKE at 100 m, the 200-20 m potential temperature gradient, and winds at 80 m and 120 m are treated as additional atmospheric variables in the sensitivity analysis. Based on the improvements observed here, all of these auxiliary variables are included for the main comparisons, though some provide limited conditional information.

Comment #2.C11

- *Discussion around Fig. 4 is slightly difficult to follow, and seems almost contradictory. Your discussion suggests the improvements are small, "indicating only limited incremental benefit", and at line 350 onwards you give further reasons why these improvements shouldn't necessarily be trusted. But you then conclude that all inputs should be retained anyway. Please clarify this discussion. Also, since you decide to include all inputs anyway, a more conventional way to perform this investigation would be to start with all inputs, and perform an ablation study to investigate the influence of each. You should justify your investigation approach.*

Authors' Response:

We have revised the text to explain more clearly the purpose of the sensitivity analysis (lines 385-387). This is a forward, one-factor-at-a-time sensitivity study relative to a simple 100 m wind-only baseline, rather than a leave-one-out ablation from the full input set. This design is chosen to quantify the incremental benefit of adding each source of atmospheric information to a physically interpretable baseline. Although the improvements from individual auxiliary variables are generally small, most are consistently beneficial and the additional wind levels provide the largest improvement. Therefore, we retain all auxiliary variables in the main model comparisons.

Comment #2.C12

- *It would be informative to experiment with reducing the temporal range of the training data, in comparison to varying the number of layout scenarios seen in training. Given your Fig. 3 results, and that the number of scenarios is several orders of magnitude smaller than the number of time snapshots, it seems likely that a better use of computational resources would be to decrease temporal coverage and increase layout diversity; this would be a useful insight to provide, if you find it to be true.*

Authors' Response:

We agree that the trade-off between temporal coverage and layout diversity is an important question. Our study attempts to mirror standard operational practice which uses full TMY in assessment. Although reducing temporal coverage is technically straightforward, it could introduce seasonal biases and reduce the diversity of atmospheric conditions represented in training. Furthermore, increasing layout diversity would require additional WRF-WFP simulations beyond the currently available scenarios.

The case study therefore does not establish the optimal balance between atmospheric state diversity and layout diversity. Determining the minimum temporal coverage and number of distinct wind farm layouts needed for robust surrogate training is an interesting direction for future work, but is beyond the scope of the present study. We have revised lines 581-584 to reflect this point.

Results:

Comment #2.C13

- *Results are mostly assessed in terms of power, rather than wake-induced losses. I suggest that results should either be expressed in terms of wake-induced losses, or compared against a baseline no-wake model. At the least, the authors should explain why the error metrics are not exaggerating the model performance.*

Authors' Response:

Thank you for this suggestion. We agree that evaluating only total power may obscure errors in the smaller wake-induced signal. To address this, we have added an analysis of wake-induced farm power loss using the shared proposed farm between Scenarios 11 and 12. We calculate the relative wake-induced loss from the difference in mean farm power between these scenarios, normalized by the standalone mean power from Scenario 11, and present the results in Section 4.2.5 (lines 480-496).

The results show that all surrogate models closely reproduce the WRF relative wake loss of 7.67%, with estimates ranging from 7.62% to 8.05%. The instantaneous error metrics further show differences in predictive performance, with DDPM achieving the lowest RMSE (8.40 MW) and higher ensemble coverage (94.140%) than DDPM-Res (72.46%).

We have also replaced the Scenario 12 power time-series plot with a comparison of wake-induced power loss, as the original plot provided limited additional information beyond the Scenario 11 results. The new results show that the models capture the smaller wake-induced loss signal despite being trained to predict power production rather than wake loss directly. The corresponding tables and figure are presented in the revised Results section (lines 480-496 and 521-526).

To further investigate the relatively large instantaneous wake loss errors, we have added an analysis in Appendix D (lines 639-653) to examine their distribution and dependence on wind conditions. The largest 1% of absolute prediction errors account for approximately 73–80% of the total squared error. About half of these extreme errors occur during rapid 10-min wind speed changes or coincide with negative wake loss events. We also examined the effect of temporal averaging and found that RMSE decreases by about 26% at hourly, 63% at 6-hour, and 81% at daily resolutions. We have further clarified the limitations of instantaneous wake loss predictions and discussed the potential benefits of incorporating historical atmospheric information to better capture rapidly changing wind conditions.

Comment #2.C14

- *Fig. 8 and surrounding discussion: this in particular would likely be more informative if investigated as wake-induced losses (or "normalized wind power" as in Section 4.2.6), rather than power. It is easy to get close to the "truth" curve armed only with a WRF simulation (no farms) and the turbine power curve. The wake-induced losses are the important output, and would show the differences between your modelling approaches much more clearly when expressed as a timeseries.*

Authors' Response:

We agree with this comment. We have replaced the Scenario 12 power time-series figure with a comparison of wake-induced power loss and added tables reporting aggregate wake losses and instantaneous prediction errors in Section 4.2.5 (lines 480-496). These results provide a more direct assessment of the models' ability to reproduce wake effects, as detailed in our previous response to Comment 13.

We have also expanded the Discussion (lines 576-579) to emphasize that predicting wake-induced losses is a more stringent test, since the loss signal is substantially smaller than total power and the models are trained to predict power production rather than wake-induced losses directly.

Discussion:

Comment #2.C15

- *The authors are correct to identify lack of domain generalisability as a key limitation. This should be made clearer near the beginning of the discussion, since this undermines the "removing a key cost barrier" claim. Given the workflow as presented in this study, the method relies on a large volume of training data being generated on the same domain as the intended application. Given the range of ongoing and future work discussed, I recommend re-framing the current study as a preliminary or proof of concept towards the stated aim of removing the computational cost barrier. The similarity between the training and evaluation scenarios should also be made clearer throughout the results and discussion, as this is a key caveat which limits the strength of the conclusions which can be drawn.*

Authors' Response:

Thank you for this important point. We agree that the original wording needed to distinguish more clearly between multi-farm deployment and cross-domain generalization. The U-Net surrogate developed here is already being used for rapid assessment within the study domain, where the initial WRF-WFP simulations provide

the training data and the trained surrogate can then evaluate additional candidate layouts at much lower computational cost. In this setting, the method addresses an important computational bottleneck associated with repeated layout evaluation. We have removed the "removing a key cost barrier" claim and revised lines 554-557 in Section 5 to reflect this.

Since the present study does not demonstrate transfer to a new geographical domain without additional high-resolution training simulations, we have moderated the broader claims about removing the computational cost barrier and clarified that the current results apply to repeated assessment within a fixed domain (lines 568-574). Cross-domain transfer is being investigated separately in our ongoing multi-domain work, where the approach has demonstrated effective generalization across domains.

Missing discussion points:

Comment #2.C16

- *Limitations of the underlying WRF-WFP model. E.g. where more than one turbine is present in a single cell, wake effects are not properly represented. It would also be of benefit to review WRF-WFP validation studies in the introduction to add credibility to the overall workflow.*

Authors' Response:

We thank the reviewer. We have added a review of WRF-WFP validation studies to the Introduction (lines 33-36). Regarding appropriateness of multiple turbines per grid cell, please refer to our response in Comment #2.C5.

Comment #2.C17

- *You mention in the results section that the models perform better under stable wind conditions. It seems likely that this is related to the use of a single instantaneous snapshot being delivered as input to the surrogate model. The limitations of this choice should be discussed.*

Authors' Response:

Thanks for raising this point. We interpret the reviewer's reference to stable wind conditions as referring to slowly-varying wind conditions. The use of instantaneous atmospheric inputs may be one factor contributing to these discrepancies, since the surrogate does not explicitly receive information about the evolution of the flow. However, the observed mismatches may also arise from other sources of surrogate error and the current analysis does not isolate the effect of temporal context.

Our additional wake loss analysis in Appendix D (lines 639-653) further shows that extreme instantaneous errors often coincide with rapid wind speed changes, while temporal averaging can reduce these errors significantly. This indicates a limitation of instantaneous predictions and suggests that incorporating historical atmospheric information may be beneficial when 10-minute wake loss predictions are required.

Importantly, the diffusion models show adaptive uncertainty during rapidly changing wind periods, where the discrepancy between the deterministic prediction and the WRF-WFP target increased, the predictive ensemble generally widens as well. Therefore, although the instantaneous inputs may limit the representation of the

evolution of the flow, the probabilistic models are still able to reflect increased uncertainty when the prediction problem becomes more difficult. We have revised the discussion to clarify this interpretation (lines 609-612).

Comment #2.C18

- *You mentioned earlier the contrasting computational cost associated with inference of the U-Net and diffusion models. This should be revisited here to inform your comparison of the benefits and weaknesses of the two approaches.*

Authors' Response:

We have expanded the discussion to make the computational trade-off between the U-Net and diffusion models more explicit (lines 569-574). We emphasize that the U-Net provides accurate point prediction and significantly lower inference cost, and is better suited to rapid screening of many candidate layouts within a fixed geographical domain. While the diffusion models require higher inference cost, this additional cost enables probabilistic predictions that are more appropriate when uncertainty quantification is important.

Technical corrections:

Comment #2.C19

- *Abstract: "Overall, the all models show..."*

Authors' Response:

This grammatical error has been fixed.

Comment #2.C20

- *Line 480: "larger when atmospheric covariate changes rapidly"*

Authors' Response:

This grammatical error has been fixed.

References

- Pryor, S.C. and Barthelmie, R.J., 2024. Wind shadows impact planning of large offshore wind farms. *Applied Energy*, 359, 122755.
- Dong, Z., Zeng, W., Deng, X. and Yang, K., 2027. Machine learning for wind farm wake modeling: Output-driven taxonomy, architectural inductive biases, and benchmarking. *Renewable and Sustainable Energy Reviews*, 243, 117359.
- Baas, P., 2024. *Winds of the North Sea in 2050—Public Final Report.*
- Dörenkämper, M., Vollmer, L. and Lakdawala, Z., 2025. Simulating the impact of offshore wind expansion on yield and efficiency in the North Sea by 2050.
- Akhtar, N., Geyer, B., Rockel, B., Sommer, P.S. and Schrum, C., 2021. Accelerating deployment of offshore wind energy alter wind climate and reduce future power generation potentials. *Scientific Reports*, 11, 11826.

Additional correction

During the revision, we identified and corrected a transcription error in Appendix Table C1 affecting the training-set results for the U-Net-CF-L_1 model under Scenarios 1 and 2. We have verified the corrected values against the original model evaluation outputs. This correction does not affect the main findings or conclusions of the study.

We also removed Appendix F to reduce redundancy, as the results presented in the main text already demonstrate the model's strong performance at the farm level.

We thank both reviewers again for their careful evaluation of the manuscript and for their constructive comments and suggestions.