Responses to Reviewers' Comments for Manuscript 10.5194/wes-2025-261

Graph Neural Operator for wind farm wake flow

Addressed Comments for Publication to

WES Wind Energy Science

by

Jens Peter Schøler, Frederik Peder Weilmann Rasmussen, Julian Quick,

and Pierre-Elouan Réthoré

Document Overview

This document contains the authors' responses to the reviewers' comments for manuscript wes-2025-261, "Graph Neural Operator for wind farm wake flow." The document is structured as follows:

- **Section 1:** Responses to Reviewer 1's comments
- **Section 2:** Responses to Reviewer 2's comments
- **Appendix:** Manuscript with tracked changes

The appendix presents a marked-up version of the revised manuscript generated using `latexdiff`, with deletions shown in red strikethrough and additions shown in blue. Please note that the figure and table formatting in the tracked changes version is suboptimal and for inspecting those the main file should be consulted.

Authors' Response to Reviewer 1

General Comments. 2025-12-12 The manuscript wes-2025-261 presents an advanced deep-learning approach for constructing a surrogate model to predict wind-speed deficits within and downstream of wind farms. The model is trained on a large data set generated by a low-fidelity, steady-state wake model (PyWake). The authors highlight prediction speed and generalisation to a range of wind-farm layouts and inflow conditions as key advantages. The deep-learning architecture is described in detail, and many test results are provided. Although the model architecture is sophisticated and the method is thoroughly described, I believe the manuscript requires major revisions for the following reasons:

Response: Thank you for your thorough review of our manuscript.

Your detailed feedback has improved the quality and clarity of the paper. We appreciate the time you invested in examining both the methodology and presentation of our work. We have addressed your major concerns, by:

- Toning down claims of "breakthrough" in favor of more measured language emphasizing our methodological contribution
- Clarifying the nature of our contribution relative to existing GNN-based approaches in the literature
- Revising the abstract to remove potentially misleading metrics, and renamed MAPE to MANE as to not use incorrect definitions
- Moving technical implementation details to the appendix to improve readability for the WES audience
- Adding new visualizations including complete flow map comparisons with error plots (new Figure 11)

We have also addressed the specific comments, making corrections to formatting, clarifying explanations, and adding requested information throughout the manuscript. Below, we provide point-by-point responses to each of your comments.

Major Comments

Comment 1

The manuscript concludes that the proposed method represents a breakthrough in data-driven wind- farm modelling and provides a robust baseline for efficient flow prediction. However, the literature already contains ML-based surrogate models, including GNN-based and other approaches. The manuscript does not demonstrate that the proposed model outperforms these (simpler) existing alternatives or offers clear additional value. To substantiate the claimed contributions, a comparative analysis is needed.

Response: We appreciate the reviewer's comment and have toned down the claim of "breakthrough" in favor of more measured language emphasizing the methodological contribution, as also suggested by Reviewer 2.

We have also tried to make the nature of our contribution clearer in the revised manuscript. The novelty lies not in outperforming existing GNN-based models on the same task, but in addressing a different prediction target: spatially continuous flow fields across the entire wind farm, rather than turbine-level quantities. As discussed in Section 1, previous GNN applications to wind farms focus on predicting power or loads at turbine locations, while other machine learning approaches has their own limitations. We have added the following text to Section 1 clarifying the advantages of the graph-based approach over alternatives such as CNNs and MLPs:

" The graph-based formulation offers distinct advantages over alternative neural network architectures for this application. Convolutional Neural Networks (CNNs) require fixed grid structures and, therefore, struggle to accommodate varying turbine layouts across farms with different sizes and densities. Simultaneously, they need a fixed grid, often with limited resolution, which imposes an upper resolution limit, whereas a GNN-based approach enables inference at the exact positions of interest. Multilayer Perceptron (MLP) based surrogate models for single-wake prediction have been proposed, but these still require a classical superposition scheme to reconstruct the full farm flow, inheriting the limitations of algebraic wake summation. In contrast, the message-passing framework inherent to GNNs naturally represents turbine-turbine interactions through graph connectivity, enabling the model to learn meaningful turbine interactions, including nonlinear interactions, that generalize across diverse layouts without relying on explicit superposition. For a review on data-driven methods in wind farms fluid flow, see e.g. Zehtabiyan-Rezaie et al. (2022). "

Comment 2

The performance metrics reported in the abstract and conclusion of the manuscript (RMSE and MAPE) are difficult to interpret (without having read in detail the complete manuscript) and, in my view, give an overly optimistic impression. As the primary purpose of PyWake—and by extension of the surrogate model—is to accurately represent wake flow velocity deficits, the MAPE should be defined with respect to the ground-truth velocity deficit, instead of to the free-stream wind speed. As shown in Figure 9, for moderate to high wind speeds, the predicted deficits in waked zones deviate by more than 100% from the ground-truth wind velocity deficits. In addition, the reported RMSE value is averaged over a large domain that may include large zones without any wind velocity deficit, where the prediction problem is trivial. This may inflate this performance metric and should be clarified or adjusted. In order to assess the accuracy of the proposed GNO approach, it may also be useful to compare its accuracy to the accuracy of PyWake itself (either relative to real-world data or high-fidelity models), as reported in previous studies.

Response: We agree that presenting a non-standard variant of MAPE in the abstract was somewhat misleading, and that the remaining metrics were also challenging to interpret without the full context of the problem. To address this, we have removed the performance metrics from the abstract. Regarding the MAPE metric specifically, we have renamed it to Mean Absolute Normalised Error (MANE) throughout the manuscript to avoid confusion with the standard definition. We have also added an explanation clarifying its relationship to conventional MAPE and why standard MAPE is problematic in this application, specifically, normalising by the target velocity deficit leads to numerical instability when deficits approach zero in lightly waked regions.

Regarding the concern that RMSE is inflated by unwaked regions that are supposedly trivial to predict, we would argue that accurate prediction in these areas is nonetheless important: a model that erroneously predicts wake deficits where none exist would be equally problematic for downstream applications. Furthermore, as illustrated in Fig. 3, the adaptive bounding box used for data generation ensures that the evaluation domain is constrained to regions of physical interest, thereby limiting the influence of far-field areas with negligible wake effects.

Finally, regarding the suggestion to compare the GNO's accuracy against PyWake's own accuracy relative to high-fidelity data or measurements: while we appreciate the value of such a comparison, we consider it outside the scope of the present work. The focus of this study is on the development and evaluation of the GNO as a surrogate, rather than on validating PyWake itself. Validation studies of PyWake have been reported elsewhere in the literature.

Comment 3

The abstract notes as motivation for implementing a surrogate model, the many required simulations for applications such as wind-farm layout optimisation and consideration of neighbouring wind farms. It is therefore recommended that the manuscript explicitly addresses whether—and by how much—the proposed surrogate model really adds value for these targeted applications, as well as any limitations that may arise from using the GNO-approach compared to using PyWake.

Response: Thank you for this comment. We acknowledge that the original abstract was misleading regarding the model's intended use for layout optimization. The abstract has been revised to emphasize that the primary motivation is assessing wake effects from neighboring farms, which requires many simulations. We now mention that such surrogates may make it easier to consider neighboring farm wakes during layout optimization, rather than suggesting the model is directly intended for WFLO. In Section 3.2, we have added a discussion of how the GNO's two-stage architecture suits the neighboring farm scenario. Specifically, a neighboring farm only needs to be encoded once, after which the latent turbine states can be reused to evaluate wake deficits at arbitrary query locations. This allows candidate positions to be assessed without reprocessing the neighbor each time, which is advantageous when exploring many layout alternatives. We believe this framing better reflects the model's current capabilities and its most promising near-term application.

Comment 4

Readers of Wind Energy Science are typically not AI specialists. The manuscript contains many low- level implementation details that obscure the core ideas and may discourage readers from completing the lecture of the manuscript. Therefore, consider moving purely technical implementation aspects to an appendix. As the software code is openly available (which is greatly appreciated), some details that are generic to deep-learning workflows might even be omitted entirely from the manuscript.

Response: Thank you for acknowledging our sharing of the code. We understand that we may have included too many details; however, we believe that precisely because WES readers are not AI specialists, they may benefit from a more thorough description of the implementation. We also believe that this is in line with the change to a stronger focus on the methodological aspects of the paper.

That said, we have moved some of the generic deep learning details to the appendix, including dropout regularization and layer normalization. The grid search tables from the results section have also been relocated to the appendix.

Specific Comments

Abstract

Comment 5

Line 3: The text implies that the classical superposition principle is a disadvantage of existing approaches, yet the surrogate model is trained by a model that implements such an approach. Please rephrase to avoid suggesting a contradiction that is only clarified in the conclusion of the manuscript.

Response: We have now added the distinction that this is something that will be valuable in the future when we train on high-fidelity data.

Comment 6

Line 8: "simulated wind farms": I recommend that, for clarity, it is mentioned explicitly that it concerns PyWake or, more generally, a low-fidelity steady-state engineering wake model.

Response: The suggestion has been added menitioning PyWake and explaining it is a low-fidelity steady-state engineering wake model.

Comment 7

Line 10: "underestimated ... wake effects". For clarity, add: "compared to the simulated values".

Response: This has been added.

1. Introduction

Comment 8

Line 61: Many WES readers will not be familiar with Graph Neural Operators (GNOs). Since GNNs are introduced earlier and better known, consider briefly describing how GNOs relate to GNNs and why they may yield improved performance.

Response: Thank you for this suggestion. We have added a brief explanation clarifying the relationship between GNOs and GNNs.

Comment 9

Line 74: Use "Section" (capital S). In general, ensure consistent capitalisation and punctuation throughout the manuscript (e.g., figure 2, line 121, line 129, line 140, line 209, line 240, line 320, line 404, line 416, ...)

Response: We have corrected the Section issue and have made the correction to the WES Copernicus style "Sect." now, except for when starting a sentence where it is referred to as "Section". We are not exactly sure what the issues around Figure referred to in the lines, qua the WES guideline "Figure" if it occurs as the first word and "Fig." inside the text. We have cleaned up some inconsistencies in this formatting now and believe it is done correctly.

2. Methodology

Comment 10

Line 78: "Here, . . . ". I suggest improving the formulation. (at multiple occasions in the manuscript)

Response: We agree that this formulation has been used much especially in the Methodology section. We have therefore rephrased Line 78 and line 275. (Some similar occurrences has been moved to the appendix, due to input of Reviewer 2)

Comment 11

Line 104: Alternative formatting: R_0^+

Response: Updated, that looks better.

Comment 12

Line 120: Should appear on line 119.

Response: Fixed.

Comment 13

Lines 121 – 126: Difficult to follow; please rephrase for clarity.

Response: The order of the information in the paragraph has been reworked, and split into two parts. It now reads:

"Correlated inflow conditions are generated using Quasi-Monte-Carlo sampling with the improved Sobol sequence (Joe and Kuo, 2008). While Dimitrov et al. (2018) employed the Halton sequence for similar purposes, we tested both approaches and found that the Sobol sequence, also used by de Santos et al. (2024), better captures extreme values in the distribution tails.

Figure 1 (e) illustrates the two-dimensional quasi-random samples from the Sobol sequence before transformation. The sampled component used to generate the free-stream velocity U is denoted p_U, and p_{TI} denotes the component used to generate the ambient TI."

Comment 14

Line 130 - : This section presents extensive detail on selected model parameters. Clarify which parameters materially affect results. Non-essential implementation details may be moved to an appendix. Additionally, because not all readers are familiar with PyWake, please summarise its key assumptions (steady-state flow, homogeneous inflow, no yaw misalignment or curtailment, etc.). Also, clarify whether the same CT- curve is used for all simulations, and discuss implications for generalisation. (As you may know, examples in literature exist attempting to integrate the CT-curve as node features in the GNN).

Response: We agree that too many details was given here and as we use default parameters this much emphasis in not necessary. We have therefore decided to move these details to the appendix, as also suggested by Reviewer 2 (Comment 1). In addition we have summarised the key assumption for unfamilar readers now.

We have added the following paragraph:

" PyWake is an open-source steady-state engineering wake modeling framework that computes wake deficits and turbine interactions. The framework supports spatially varying inflow conditions, multiple turbine types, and yaw misalignment via wake deflection models. However, to establish a baseline dataset, several simplifications were adopted: (i) homogeneous inflow with uniform free stream velocity and turbulence intensity; (ii) a single turbine type (DTU-10-MW) with a fixed C_{T}-curve; and (iii) all turbines aligned with the inflow direction. These choices reduce the input parameter space but may limit applicability to scenarios involving heterogeneous inflows, mixed turbine fleets, or active wake control."

Comment 15

Line 133: Citation formatting issue.

Response: We have split the citation to be explicit about its interpretaion. It now reads: The wakes are modeled with the updated self-similar Gaussian single wake deficit model NiayifarGaussianDeficit by Niayifar and Porté-Agel (2016), which is a further development of the Gaussian wake model by Bastankhah and Porté-Agel (2014).

Comment 16

Line 133: "This model was chosen due to its relative simplicity". Does this have any influence on the generalisation of the results of the manuscript? Please make this explicit in the manuscript.

Response: The sentence was changed to highlight the TI dependence rather than simplicity. "This model was chosen for its TI-dependent wake expansion to ensure a coupling between TI and velocity deficits."

We have also added the following text after introducing the numerical schemes, to better capture the nature of the design choices for the dataset generation, the changed also address comment 20:

"However, the use of All2AllIterative also benefits the GNO application: Because turbine interactions are resolved iteratively in both directions, the resulting input-output mapping constitutes a non-linear operator, providing a more challenging and representative test case for the GNO than a simple PropagateDownwind scheme. Although PyWake employs linear wake superposition (Eq. 2), the coupled iterative solution introduces non-linearity that the GNO must learn to approximate. More broadly, the dataset was designed to capture diverse physical effects, including blockage, turbulence-dependent wake expansion, and bidirectional interactions, rather than to maximize fidelity to any single high-accuracy model. Such diversity ensures that the GNO learns a sufficiently complex operator, demonstrating its capacity to generalize across varied flow physics. Fu- ture work should investigate the impact of both different superposition methodologies and higher-fidelity training data on GNO performance."

Comment 17

Line 135: "The authors fitted the model to a LES." Why is this relevant here?

Response: This has been reformulated and moved to the appendix as additional details for interested readers.

Comment 18

Line 143: "…model…chosen for its simplicity". Does this have any influence on the generalisation of the results of the manuscript? Please make this explicit in the manuscript.

Response: The direct reference to being chosen for simplicity has been removed as it is not fully accurate. With regards to generalisation see reply to comment 16

Comment 19

Line 145: When "TI" is used as a variable, the formatting shall be that of a variable. (This comment is applicable throughout the manuscript.)

Response: Thank you for the comment.

We have updated TI in equations and as a variable to be I_0 for ambient TI and I_a for added TI.

Comment 20

Line 170: "A linear sum is used for the wake summations." Here, it could be stated/repeated that the GNO theoretically may also handle non-linear superposition, but this may be verified in future work.

Response: We have updated the explanation of the use of linear-superposition slightly further down in the text now, to clarify some of this and acknowledge that further studies are needed. See reply to comment 16

Comment 21

Line 179: "The GNO is gird invariant." Begin a new paragraph.

Response: We have reworked the sentence to start a new paragraph here.

Comment 22

Line 180: "...from flow away from..." ?

Response: This was addressed in rewrite mentioned in comment 21.

Comment 23

Line 186: $\bar{x}$ could be interpreted as a vector or maximum; consider using: x'

Response: We agree that the overline was an unfortunate choice, but we have reserved the use of the apostrophe for neural network outputs. Therefore, we have instead changed to using $\widetilde{x}$ which is more commonly used for coordinate transformations than the original overline.

Comment 24

Figure 3: Clarify why the coordinate origin differs from the farm centre. How the coordinate origin is determined? Does this influence (the features of) the GNO?

Response: The layouts are created with the PlayGen and are not necessarily centered during generation, hence neither will the flow maps be. It doesn't change anything because the GNO is only concerned with the relative internal distances between nodes. The distance formulation has been clarified in response to Reviewer 2's comment 3.

Additionally a sentence has been added to the describe that it is a non issue: "As an artifact of the PlayGen generation methods, the wind farm center and coordinate center do not always coincide. Because the GNO uses relative internal coordinates, this does not affect its performance. A sentence explaining this has been added and more details are given in Sect. 2.2"

Comment 25

Lines 191- 193: Appear irrelevant; consider removing.

Response: We believe most readers of WES will not be familiar with graph theory graphs and therefore we believe having a bit of historical background makes the introduction slightly more elegant and engaging for the reader.

Comment 26

Line 198: "Instead, they are copied to each node as node features." Why did you make this decision? provide references if applicable.

Response: We have added the source of inspiration, to this sentence.

Comment 27

Line 206: Consider adding $V = V_{\mathrm{wt}} \cup V_{\mathrm{p}}$

Response: Added, thank you for the suggestion.

Comment 28

Line 209: Consider adding the analogous equation to Eq. 7.

Response: Added, thank you for the suggestion.

Comment 29

Line 209: "The inter-turbine edges...": start a new paragraph for readability.

Response: We have added a new paragraph.

Comment 30

Line 204: "Probe node...": start a new paragraph for readability.

Response: We believe this was meant to refer to line 215 not 204. We have made the change based on that assumption while also addressing comment 31. So that it now reads:

" Probe node connectivity is straightforward because each probe connects to all wind turbines. Inter-turbine connectivity is more complex because it requires an algorithmic approach to determine which turbines should be connected. Several algorithms exist for constructing the connectivity of the turbine nodes. Duthé et al. (2023) investigated four graph connectivity schemes: Delaunay triangulation (Delaunay, 1934; O'Rourke, 1988; Fey and Lenssen, 2019), K-Nearest Neighbors (KNN), a radius- based method, and a fully connected scheme. They concluded that Delaunay triangulation provides the best balance between accuracy and computational performance. Accordingly, we also adopt Delaunay triangulation in this work to derive inter-turbine connections. "

Comment 31

Line 216: "... it uses Delaunay triangulation..." This is repeated in line 220. Consider rephrasing.

Response: Addressed in reply for comment 30.

Comment 32

Line 224: You state that Cartesian coordinates are used as edge features (and not as node features, which intuitively would make more sense), which seems to contradict later text. Please clarify and justify the design choice for the node and edge features; add references if similar implementations exist. Also specify how the coordinate origin is chosen, as it has an influence on the Cartesian coordinates.

Response: The concerns addressed in this comment has been addressed in replies to comments 24, 26 and reviewer 2 comment 3.

The main summary being we had not clearly managed to describe the actual behavior of the model and that we now have put more emphasis into accurately describing that the GNO uses relative positions between nodes, and that the center of the graph is therefore not of importance.

Comment 33

Lines 230 – 235: The sequence and interaction of the models are unclear; Consider rephrasing.

Response: We rephrased it to first introduce the GNO and then introduce the developments of the GNO.

Comment 34

Line 239: Clarify "common" and "abstraction".

Response: This has been rephrased to:

"The encoder-approximator-decoder configuration is a widely adopted architectural pattern in GNNs, providing a modular framework for transforming input graphs into predictions (Battaglia et al., 2018)"

Comment 35

Line 240: "initial two stages": Clarify whether these are the encoder and approximator?

Response: It is, it has been updated to: "the encoder and approximator stages"

Comment 36

Line 245: "sped up"

Response: We have interpreted this as preferring a more formal language. The sentence has been changed to:

"This means that predicting in a flow field incurs lower computational cost than a fully integrated prediction scenario."

Comment 37

Line 248: "Fore instance, in fully couple formulations, ...": is this the case in your implementation?

Response: Indeed, it is. We believe the confusion around this has been addressed in the reply to comment 16, outlining the meaning of the All2AllIterative scheme.

Comment 38

Lines 256 – 260: This information deserves a separate subsection somewhere else in the manuscript.

Response: We have moved it into Section 2.4 Training and evaluation as its own subsection.

Comment 39

Line 267: "Q": here a capital letter, in contrast to in line 278.

Response: Thank you for noticing, we have updated line 278 to use all capital letters.

Comment 40

Line 278: The equation comes back (and better) in lines 292 – 294

Response: We have corrected the one in line 278, see previous reply. The reason they come back is that in line 278 it is intended to introduce MLPs in general, and the next time we see the actual dimensions, the considered MLPs transform.

Comment 41

Equation (10a): k = 1,...,K; Move be on the same line.

Response: This is done in preparation for the final format using two columns, in which there is not enough space for the second part.

Comment 42

Lines 292-294: For clarity, repeat here the numeric values in your implementation for |E|, fe, ..., and Q.

Response: We agree keeping track of all the dimension are quite difficult. We have added the following:

"For convenience, the key dimensions are restated here: the cardinalities $|V|$ and $|E|$ vary with the wind farm configuration; $f_v = 2$ and $f_e = 3$ are the initial node and edge feature dimensions; $K = 9$ is the number of RBF kernels; and Q is the latent-space dimension."

Comment 43

Line 300 and equation 12b: Shouldn't this be $\phi_{h'}$? (ref. notation in Figure 5)

Response: You are correct, there is a mistake here. The issue was Eq. 12b it has been corrected to $\phi_{h'}^{(m)}$

Comment 44

Line 309: Define $\widehat{x}_j$

Response: Thank you for the comment.

A more thorough and consistent introduction of this variable has now been implemented; it has also been renamed $\widehat{x}_q$ as to better fit the established Q for latent-space dimension:

" In this work, Softmax aggregation is used as $\rho_{e\to v}$, it uses Softmax to scale the latent space of each feature dimension independently. The incoming messages to node i from its $|\mathcal{N}(i)|$ neighbors form a matrix of shape $|\mathcal{N}(i)| \times Q$. For each feature dimension, let $\hat{\boldsymbol{x}}_q \in \mathbb{R}^{|\mathcal{N}(i)|}$ denote the vector containing the q-th feature value across all neighbors.

$$\rho_{e\to v} = \sum_{\hat{\boldsymbol{x}}_q \in \mathcal{X}} \frac{\exp(\hat{\boldsymbol{x}}_q)}{\sum_{\hat{\boldsymbol{x}}_r \in \mathcal{X}} \exp(\hat{\boldsymbol{x}}_r)} \cdot \hat{\boldsymbol{x}}_q \tag{1}$$

where $\mathcal{X} = \{\hat{\boldsymbol{x}}_1, \ldots, \hat{\boldsymbol{x}}_Q\}$ is a collection of neighborhood features to be aggregated independently. "

Comment 45

Figure 7: Nothing appears marked yellow.

Response: The wind turbine nodes after the Approximator ($\mathcal{A}$) are marked in yellow. We have added an additional visual queue at the center to make it look slightly more like a rotor/nacelle assembly.

Comment 46

Equation 18b: To be a relevant performance measure for accuracy, the MAPE of the predicted wind speed deficit should compare the predicted wind speed deficit relative to the ground-truth wind speed deficit (instead of comparing to the free inflow wind speed).

Response: As discussed in the response to comment 2 MAPE has limitations, but we agree that referring to the metric we used as MAPE is incorrect and have changed the naming to MANE.

Comment 47

Line 371: "limiting factor has been the 72-hour walltime": Was the performance the GNO still improving at the end of this walltime? If yes, why could the model not be saved to resume the training of the model, starting a new walltime?

Response: Indeed the model were converged, this is not relevant information and have therefore been removed.

Comment 48

Line 385: Grammar issue.

Response: "Is" corrected to "are"

Comment 49

Line 397: Please explain more clearly.

Response: The following explanation has been added:

"Where the term $(1 + f_e)$ accounts for the fact that each edge stores both its f_e features and an index pair identifying the connected nodes in the adjacency list representation."

Comment 50

Line 400: "To evaluate the model under realistic conditions." This suggests that the other layouts would not be realistic. Consider rephrasing.

Response: Indeed, thank you for pointing that out. We have changed it: "To evaluate the model on an established benchmark, the IEA Wind 740-10-MW reference wind farm is employed."

Comment 51

Line 401: "farm layout"

Response: Corrected to farm layout.

Comment 52

Line 405: "simple operation": Does the applied power curve not depend on TI? If yes, how is TI at the turbine locations predicted by the GNO?

Response: No, the CT curves in PyWake do not include TI dependence by default. TI affects the wake expansion through the wake deficit model, which augments the resulting velocity deficit and ultimately the effective windspeed used to determine the power and CT state of turbines operating in wakes.

3. Results and Discussion

Comment 53

Line 419: "internal MLP": Clarify what is meant by "internal".

Response: The following has been added for clarification:

"For the model configuration, the encoder, wind turbine processor, and probe processor MLPs are collectively referred to as the internal MLPs."

Comment 54

Line 229: "probes per graph": how are these selected? Randomly in the large box around the wind farm?

Response: During training, probe locations are selected using uniform random sampling across the bounding box surrounding each wind farm. This choice was made to ensure broad coverage of the domain without introducing additional hyperparameters. The revised manuscript now clarifies this in the methodology section. Please also see our response to Reviewer 2, Comment 10, for further details.

Comment 55

Line 429: "number of batched graphs": Do you mean, the number of graphs per batch?

Response: Yes, it has been corrected

Comment 56

Line 436: "increases the confidence that the right model is chosen": Why?

Response: MSE and MAE weight errors differently: MSE penalizes larger errors more heavily due to the squaring, while MAE treats all errors linearly. When both metrics rank models similarly, it suggests that the best-performing model is not merely minimizing a few large errors at the expense of many smaller ones (or vice versa), but is performing consistently well across the error distribution. This consistency across different error weightings increases confidence that the model selection is robust and not an artifact of the chosen metric.

The following text has been added as an explanation: "Since these metrics weigh errors differently, this increases confidence that the model selection is robust and not overly sensitive to the choice of metric."

Comment 57

Table 2: "lowest best metrics marked with b": What is meant by that?

Response: This was a typo it has been corrected to bold, although this table has been moved to the Appendix.

Comment 58

Table 2: Add the unit of each performance indicator.

Response: We have added the units for the performance indicators, although this table has been moved to the Appendix.

Comment 59

Table 3: Add the unit of each performance indicator.

Response: We have added the units for the performance indicators.

Comment 60

Table 3: The values of the performance indicators for the test set seem to be of another order of magnitude compared to those of the validation set. How is that possible?

Response: We had accidentally forgotten to unscale the metrics relating to the training; hence, the originally reported metrics were dimensionless for the training metrics and had units for the test metrics. This has been corrected, and now all the metrics are in the correct unscaled form.

Comment 61

Table 3: Consider adding baseline metrics for naïve deficit profile(s) (such as the zero-deficit profile), so that the magnitude of the metrics for the GNO models can be better interpreted.

Response: We have added a "naive" baseline now. Thank you for the suggestion, you are right, this makes the metrics more meaningful in a general sense.

Comment 62

Line 455: Include a wake profile(s) inside the farm (i.e., for $\bar{x} < 0$), so that not only the far-wake profile is analysed, but also the near-wake modelling is analysed.

Response: A new figure (Figure 12) has been added showing full flow map comparisons between GNO and PyWake, including the near-wake region within the farm. This figure also includes error plots to highlight where the model falls short. See also our response to Reviewer 2, Comment 11.

Comment 63

Line 461: Similar to one of the main comments: What is the range of the ground-truth velocity deficits relative to U? This may also be small.

Response: Thank you for this clarification request. The 2-5% error range reported in the original manuscript refers to errors in effective wind speed relative to U, not errors in the velocity deficit itself, as you correctly have observed.

Using the target deficit $\Delta\widehat{u}$ as a reference has proved problematic because it can be zero, leading to an explosion in the relative error. However, this is not the case for the relative error with respect to the target effective wind speed $\widehat{u}$, because the velocities of Fig. 9 (now 8) are not inside the farm area where the non-physical zero effective wind speeds can occur. Hence, we have updated the text to use $\widehat{u}$ relative errors instead.

Please note that the farms considered in this Figure in the revised manuscript have changed from the original submission, as a new set of wind farm examples was selected to improve the visual clarity of Figure 9 (see response to Comment 13 from Reviewer 2).

Comment 64

Figure 9: The plots are too small to be readable on a printed version of the manuscript.

Response: We have made significant changes to improve the visual clarity of Figure 9 (see response to Comment 13 from Reviewer 2).

Comment 65

Figure 9: Comparing the close results of dashed and full lines with different colours (and thus different wind speeds), at first view, one could think that there has been a failure with the allocation of the colours.

Response: We believe the changes to Figure 9 has addressed these issues (see response to Comment 13 from Reviewer 2)

Comment 66

Figure 10: It would be interesting to see the plots for the 3 wind speeds separately, or for the wind speed 12ms-1 only (i.e. wind speed with the highest wake losses), in case of lack of space (or in the appendix).

Response: We have added the requested figure in a new Appendix F now.

Comment 67

Figure 11: It looks as if the orientation of the wake is rotated from the main axes of the farm layouts in Figure 8.

Response: Thank you for pointing this out, you are correct we had incorrectly performed the rotational coordinate transformation for creating the Power rose. This has been corrected and to make the connection between the two plots clearer we have added a North pointing arrow to Figure 8.

Comment 68

Figure 11: Make plots larger to be readable in printed A4 version.

Response: We have updated the layout of Figure 11 to make the power roses more readable. The narrow nature of the figure is in preparation for the two column format of WES.

Comment 69

Line 481: "conditions similar to their training data": Are you suggesting that the optimised layout is less similar to your training data? If that were known, it would be recommendable to add better layouts to your training data.

Response: Thank you for this observation. This is an implicit limitation in our training data distribution. We have added the following explanation and mentioned it as an idea for further work.

"The procedurally generated training layouts do not include simple regular-grid configurations, which may explain the slightly lower accuracy on the regular IEA layout. Including such layouts in future datasets could improve generalization to such farms."

Comment 70

Line 485: In order to validate whether the accuracy of the proposed method is sufficient for optimisation applications (which is the stated motivation for implementing this GNO model), it would be very interesting to compare the AEP predictions made on the basis of PyWake and the GNO, for both wind farm layouts.

Response: Thank you for the suggestion, as discussed in our response to Comment 3, the manuscript has been revised to emphasize neighboring farm wake assessment rather than direct layout optimization as the primary application. However, we have added the requested information on the plot for interested readers.

Comment 71

Line 485: For (farm layout) optimisation problems, the absolute value of the wake losses may be less important than the relative difference between for different layouts. Therefore, consider adding measures about the relative wake power losses predicted by the GNO.

Response: See response to comment 70.

Comment 72

Line 530: I understand your choice to compare computational cost on a one-core CPU. Nevertheless, as currently the GNO can be run on 32 cores simultaneously and on a GPU, in contrast to PyWake, in my opinion, you may also add this comparison in the manuscript. Indeed, that would be the reality for someone who may have to choose between PyWake and the surrogate model.

Response: Thank you for the suggestion. However, we have chosen to retain the current comparison for the following reasons: First, the reported timings are expressed in CPU hours, which inherently accounts for the degree of parallelization, providing a hardware-agnostic measure of efficiency. Second, implementing a fair multi-CPU comparison for PyWake would require careful consideration of parallelization strategies (e.g., parallelizing across wind directions, layouts, or grid points). Without confidence that such an implementation is optimal, we risk misrepresenting the performance of PyWake. We believe it is important to be fair to its authors and avoid any comparison that could inadvertently understate its capabilities. The trained model weights are publicly available, enabling readers to evaluate performance on their own hardware configurations.

4. Conclusion

Comment 73

Line 553: "fresh". Consider alternative wording.

Response: Agreed, changed to "offers a new perspective"

Comment 74

Line 555: ‘afterthought", "established literature": What is meant here?

Response: This was slightly unclear, it was supposed be in relation to single wake surrogates. This has been rephrased to:

"It has been established as a novel approach inspired by classic engineering models, demonstrating how the superposition principle can be integrated directly into the learning process, rather than relying on algebraic wake superposition as single-wake surrogate models require."

Comment 75

Line 559: "In section 3": leave away

Response: Rehrased to "The GNO was evaluated, . . . "

Comment 76

Line 562: "The GNO is less accurate that PyWake." You did not prove that, but it is trivial. See main comments.

Response: We believe we have addressed the concern in the replies to the main comments.

Appendix A

Comment 77

Line 601: "cf. branch input, trunk input": Please clarify these terms.

Response: The context for the GNO has been added and litterature for a broader context has been suggested:

"As with other neural operators, the data comes in triplets: branch input (the wind farm graph encoding turbine positions and inflow conditions), trunk input (the probe graph encoding query locations), and target output. For a broader context, see e.g. (Lu et al., 2021; Seidman et al., 2022)."

Concluding Response

We thank Reviewer 1 for the comprehensive feedback. The review identified several important areas for improvement, and we believe the revised manuscript is stronger as a result. We have endeavored to address every comment and hope the revisions meet your expectations. We remain open to further suggestions if any concerns persist.

Authors' Response to Reviewer 2

General Comments. The authors introduce a Graph Neural Operator architecture for wind farm wake flow prediction, where turbine-turbine interactions are learned through message-passing on a graph and flow velocities can be queried at arbitrary probe locations. Trained on ~21,000 PyWake simulations spanning diverse procedurally-generated layouts, the model generalizes across farm configurations and seems to achieve low RMSE in the predictions while being approximately 10x faster than the engineering baseline (PyWake). In general, I find that the manuscript is well written and comprehensive: all components of the framework, from procedural dataset generation to GNO architecture and training, are described in sufficient detail to enable independent implementation. The breakdown of the results is also very detailed, and I appreciate the thorough reporting on computational aspects. I recommend minor revisions, with specific comments below.

Thank you for your assessment. We believe the comments are fair and constructive and have sought to address them accordingly.

Section 2

Comment 1

A lot of detail is provided for the models that are used by PyWake to generate the training data. I think some of this can be moved to the appendix.

Response: We have moved velocity, TI and Blockage model definitions to the appendix as many WES readers will already be familiar witht these. Addionally an explanation has been added to explain the use of All2AllIterative and its implication for the GNO

Comment 2

The part about the additional coordinate system is a bit confusing and could do with some extra clarification.

Response: We have simplified the explanation and reformulated it as: "Figure 3 illustrates the bounding box of a wind farm. A farm downstream axis, $\widetilde{x}$, is introduced to measure the distance behind the most downstream turbine, making it easier to compare results across different wind farm layouts". In response to Reviewer 1's comment 24, the coordinate has also been changed to $\widetilde{x}$

Comment 3

Why did you choose to use the coordinates of the sender nodes as an edge feature, instead of the relative coordinates (vector coordinates) between the the two connected nodes, as is typically done in other GNN work?

Response: We do use relative coordinates, we have just failed to report it correctly. The following correction has been added: The edges store the relative node positions (x_{ij} and y_{ij}), and Euclidean distance (d_{ij}) between connected nodes:

$$(x_{ij}, y_{ij}, d_{ij}) = \left(x_j - x_i,\ y_j - y_i,\ \sqrt{x_{ij}^2 + y_{ij}^2}\right)$$

Comment 4

I'm not sure I see the advantage of having the decoding stage setup as it is. I get that you want to first freeze the turbine node latents, so that you have the option of adding any amount of probe nodes after, but why not have undirected probe edges between a probe and other probes, as well as between a probe and the turbine nodes. I guess the current setup makes sense when only using a single probe message-passing step, but why not have multiple steps during this phase? It would increase performance (albeit with extra computational costs). Please consider clarifying this.

Response: The current design is inspired by classical engineering wake models, where the flow field is reconstructed by superposing individual turbine wake contributions without propagation between field points. Probe nodes serve an analogous role: they aggregate wake effects from all turbines but do not exchange information among themselves, as they are not wake-generating entities.

We agree that introducing probe-to-probe message passing could potentially capture additional spatial correlations. However, this would likely require higher-fidelity training data (e.g., LES or RANS) to be meaningful. This has been noted as future work in the revised manuscript.

Comment 5

Could you have not resumed training after 72 hours? Or were the models sufficiently trained by this time and it was not needed?

Response: Indeed the model were converged, this is not relevant information and have therefore been removed.

Comment 6

I'm not sure the MAPE metric is correct, as typically we would divide the error by the target value. In some sense, what you show here is more of a MAE that is normalized by the inflow velocity, which may still be a useful metric.

Response: You are correct, the deviation from the standard was actually chosen delibarately but has been the cause of much confusion. Therefore we have chosen to rename our version to MANE and the following has been added to the metrics section explaining the choice.

"While most metrics are standard, MANE is a custom variation of Mean Absolute Percentage Error (MAPE). Standard MAPE normalizes by the target value, which becomes problematic when targets approach zero as can occur near heavily waked turbines where wake superposition overestimates the deficit. Normalizing by the wake deficit instead would cause similar issues in unwaked regions. To avoid this, MANE normalizes by the free-stream velocity, providing a more robust metric that still enables comparison across different inflow conditions."

Comment 7

You mention the cardinality of the graph, is this actually used somewhere?

Response: Yes it is used when the different components of the GNO are introduced and the dimensionality of the inputs and outputs are defined e.g. for the RBF functions

$$\varphi_e \;:\; \mathbb{R}^{|E|\times f_e} \;\rightarrow\; \mathbb{R}^{|V|\times f_e K}$$

Section 3

Comment 8

You could consider moving some of the parameter tables for the hyperparameter study to the appendix.

Response: We agree that the information was slightly overwhelming and have moved Tables (1) and (2) into the Appendix to become Tables (D1) and (D2). Instead, a new Table (1) has been introduced containing just the information of the 5 best models.

Comment 9

No ablation on the number of message-passing steps (M=3 throughout) is performed. This seems like an important hyperparameter to study, given that message-passing propagates information only to immediate neighbors. Would increasing the number of steps increase performance for cluster farm setups, where wake superposition seems to be the biggest hurdle?

Response: We did some intial experimentation prior to our grid searches and the impact of amount of message passing steps was not found to change the results significantly in our setup. This could be due to many factors, and this should be investigated more thoroughly in the future. We have added the following in the manuscript:

"The number of message-passing steps (M) was fixed at 3 throughout the grid search rather than treated as a tunable hyperparameter. Initial testing on a simpler model showed low sensitivity to additional message-passing steps, suggesting diminishing returns beyond $M = 3$. While increasing M could theoretically improve performance in densely clustered farm configurations by propagating wake interactions across more neighbors, preliminary experiments did not support this hypothesis. However, the optimal value of M is suspected to be data-dependent; higher-fidelity datasets capturing more complex wake dynamics or turbulence interactions may benefit from additional message-passing steps to fully resolve inter-turbine dependencies. A systematic sensitivity study on M across diverse farm configurations and data fidelities should be considered for future work."

Comment 10

The training procedure regarding probe node sampling deserves clarification. How are the probe locations selected? Random uniform sampling across the domain, or weighted toward regions with stronger wake effects? This choice likely influences model performance in different flow regions.

Response: For the models documented in the manuscript the probe node sampling during training is strictly random uniform sampling. An explanation has been added to the manuscript in the methodology section about training.

Here included as bonus information, early in the process we experimented with some weighed sampling. However, the amount of parameters became too overwhelming and some simplifications had to be made and this was one of them. However, if someone should be interested in implementing this we were considering this weighting for the domain (although some of the parameters should be updated slightly):

$$p_{\text{linear}}(x) = \begin{cases} 1 + k \cdot (x - 20D) & x < 20D \\ 1 & 20D \leq x \end{cases} \quad , \qquad k = \frac{1}{x_{\max} - x_{\min}}$$

$$p_{\text{cross}}(y) = 1 - \tanh\left(k \cdot y^2\right) \quad , \qquad k = \frac{2}{(y_{\max} - y_{\min})^2}$$

$$p_{\text{circ}}(x, y) = 1 - \frac{1}{1 + k \cdot ((x - x_{\min})^2 + (y - y_{\min})^2)} \quad , \qquad k = \frac{100}{(x_{\max} - x_{\min})^2 + (y_{\max} - y_{\min})^2}$$

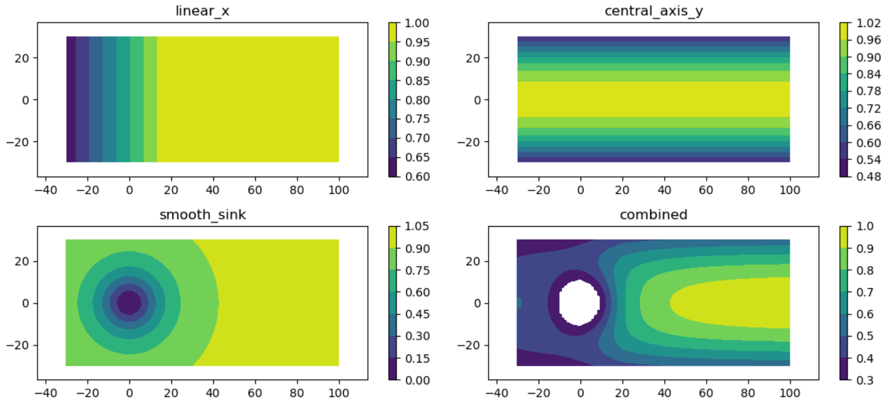

Figure 1: Sampling weights (p)

$$p(x,y) = \begin{cases} 0.3 & p_\sigma < 0.3 \\ p_\sigma & 0.3 < p_\sigma < 1 \\ 1 & 1 \leq p_\sigma \end{cases} \quad , \qquad p_\sigma(x,y) = p_{\text{linear}}(x) \cdot p_{\text{cross}}(y) \cdot p_{\text{circ}}(x,y)$$

Comment 11

I would have really liked to see a complete flow map comparison between the GNO's output vs PyWake (along with a difference plot), especially in the near-wake region. It would be extremely useful to understand where the model falls short. You could do this for the IEA Wind 740-10-MW reference reference farms for instance.

Response: We have now added a new Figure 11, that demonstrates the limitations of the model and have added some additional text.

Comment 12

This point is optional, but I was wondering what the learned RBF kernels look like. Showing the learned RBF distributions after training would provide insight into what spatial scales the model finds important.

Response: We have moved the existing RBF figure to the results and extended it with the trained Kernel and overview of the kernel values. And a short explanation of whats in the figure. The figure is numbered 13 now.

Comment 13

Figure 9. I think that you can plot the farm in a nicer way, it is hard to actually see the positions of the turbines.

Response: We have improved Fig. 9 (now Fig. 8) significantly, the graphs has been made more visible by removing empty space in the plots. This has also been applied in the velocity deficit plots. In the process a new set of graphs was chosen randomly. The text and Fig. 10 (Now Fig. 9) has been updated accordingly.

Comment 14

Figure 12. In general I find this plot a bit hard to read, the color scheme can be improved and the bars made a bit wider. For subplot (a): the KDE extends into negative RMSE values, which is not physically meaningful. Consider using a boundary-corrected KDE method, or simply truncating/reflecting at zero.

Response: We have decided to make this a two column format figure now to enable wider bars, we have also changed the order of appearance slightly as to make better use of the new format. With regards to the colors we have made a change for better visual assesment while still avoiding red-green colorblindness issues. Your assesment of the KDE was correct, and we have decided to forego it and instead use a simpler binned error count and show the count instead. The KDE has also been removed from the methodology section.

Section 4

Comment 15

I would suggest more measured language that emphasizes the nice methodological contribution without overclaiming impact ('breakthrough').

Response: Agreed we have toned down the impact and adopted an assessment more in line with the methodological contribution as suggested. We also changed a line in the abstract mentioning this.

Minor Technical Comments

Comment 16

Line 120, line 283: alignment issues

Response: We have fixed the issue.

Comment 17

Line 320: text overflow

Response: We have fixed the issue.

Comment 18

“Windfarm“ vs “Wind farm“ used interchangeably, try to stay consistent. Same for “flowmap“ vs “flow map“.

Response: We have alligned these to be consisitent now choosing two words.

Comment 19

Table 2: typo in caption (’b’ instead of ’bold’?)

Response: We have fixed the issue.

Concluding Response. We are grateful for your positive assessment of the manuscript and your constructive suggestions for improvement. Your recommendations have strengthened the presentation of our results. We also appreciate your suggestion to adopt more measured language regarding our contributions, which has helped better frame the methodological nature of this work. We have been able to accommodate your suggestions and hope the revised manuscript meets your expectations.