Date July 17, 2026
Contact person Daan van der Hoek

E-mail D.C.vanderHoek@tudelft.nl
Subject Response to reviewers

**Delft University of Technology**

Delft Center for Systems and Control

Address
Mekelweg 2 (ME building)
2628 CD Delft
The Netherlands

www.dcsc.tudelft.nl

Reviewer #1, Reviewer #2
*Wind Energy Science*

Dear Reviewers,

We greatly appreciate all the constructive comments that we received on the paper. We are confident that these comments have significantly improved the quality of the paper. In response to your feedback, we have made several revisions to the original manuscript. In this document, we aim to address all points raised by the reviewers. These responses are highlighted in blue. All changes made to the paper are also included in red. An additional document has been attached to our response, highlighting all the changes made to the paper.

Yours sincerely,

Daan van der Hoek
Tim Dammann
Jan-Willem van Wingerden

Enclosure(s): Response to comments of Reviewer #1
Response to comments of Reviewer #2

# Response to comments of Reviewer #1

Dear Authors,

I very much enjoyed reading and reviewing your paper. I think it is a very good piece of work, worthy of publication in the Wind Energy Science Journal after minor revisions. If the major and minor comments listed below are addressed, I support a publication in WES.

Best regards,

Reviewer 1

Major Comments:

1. Your GP surrogate is trained on inflow slices from a single wake. Have you validated that this is sufficient to represent wind farm configurations, where multiple wakes are typically present and interact? While I understand the useful computational efficiency of your approach, please comment more on the validity to go from single wake to wind farms, and state this as one of the limitation of your study. A validation against wind farm cases (with multiple wakes) would be very interesting to see (I am not asking for it in this paper, but you can state as required future work for improved validation).
   The sector-averaged inflow representation is layout-agnostic by design: the surrogate model takes local rotor-plane velocity and turbulence intensity as inputs, regardless of how many upstream turbines contributed to those conditions. The 3-turbine array simulations included in this study provide a first validation step, as the flow slices from the second and third turbine capture the effect of accumulated wake turbulence. Nevertheless, we acknowledge that more complex multi-wake interactions (e.g., partial overlap of multiple wakes, wake-added turbulence from non-aligned turbines) may not be fully represented and have added a sentence in Section 2.4 acknowledging this as a limitation and topic for future validation:
   "From these simulations, flow slices from the wake of the second and third turbine were collected. For now, we assume that these inflow conditions are representative for larger wind farm configurations, but this requires additional validation at a later stage."

2. At many occasions, you use symbols without properly defining them. Around an equation or any results, all used symbols should be clearly defined in the text, without ambiguity.
   Thank you for raising this point. We have systematically revised all equations and results sections to ensure every symbol is defined on first use. In particular, we resolved three symbol conflicts present in the original manuscript: (1) the atmospheric lapse rate, previously denoted $\gamma$, has been renamed $\Gamma$ to avoid confusion with the yaw misalignment angle $\gamma$; (2) the azimuth angle $\theta$ in Eq. (6) has been changed to $\psi$ to avoid confusion with the potential temperature $\theta$; and (3) the radial position of the pitch bearing in Eq. (9), previously denoted $\phi$, has been renamed $\chi$ to avoid confusion with the wind direction angle $\phi$.

3. In addition to text clarifications of symbols, it would be good to add a complete "Nomenclature and abbreviations" section in the appendix, for clarity of all your symbols and acronyms along the paper.
   An appendix with nomenclature and abbreviations has been added to the paper as suggested by the reviewer.

4. Fig. 8: It is highly unusual to see negative values for the Strouhal number. In standard aerodynamic, the Strouhal number is defined as a strictly positive quantity. I understand the sign convention that you use in Eq. (1) to separate CW and CCW Helix cases, but this actually applies on your actuated $f_{Beta}$, not on the Strouhal itself. In Figure 8, I would replace the x-axis with $(f_{Beta} - f_r) * D/U_{inf}$. Also in the caption of Figure 8, add clear separation: "comparing the CW $(< 0)$ and CCW $(> 0)$ helix implementations". Looking at the figure alone, it is not clear where are located CW and CCW for a non-very careful reader (you can even think of adding CW and CCW above the first plot where it is applicable, if you want your plot to have a clear impact).
   Thank you for this suggestion. The new label for the x-axis makes the distinction between the CW and CCW implementations much clearer. We have replaced the Strouhal number with the expression suggested by the reviewer. Furthermore, the figure now also indicates the CW and CCW regions in both the figure itself and the caption.

5. Fig. (9) and Table (4) present results of the GP performances over the whole aggregated dataset (including cases of greedy control, yaw misalignments, and Helix operation), right? It would be very interesting to see how the performances of the surrogate change whether the turbine is being actuated with WFC, or not (so de-aggregate the RMSPE results into cases of greedy, wake steering and Helix-actuated turbine). Then the robustness of the surrogate approach for WFC could better be evaluated. If the results in Fig. (9) and Table (4) include solely greedy-operating turbine in the wake of an upstream actuated turbine, this should be clearly clarified (I am unsure of this information).
   The validation results in Fig. 9 and Table 4 include all control cases: greedy control, wake steering, and helix actuation. To make this explicit and to address the reviewer's request to evaluate surrogate robustness per control strategy, we have updated Fig. 9 to distinguish the different control strategies by color. No systematic bias is observed for any particular strategy, demonstrating that the surrogate model generalizes well across all operating conditions in the database. A clarifying sentence has been added to the manuscript:
   "To assess whether prediction accuracy varies across control strategies, Fig. 9 distinguishes between baseline, helix, and wake steering cases. No systematic bias is observed for any particular strategy, confirming that the surrogate model generalizes well across all operating conditions included in the training data."
6. Section 3: After reading the whole paper, it is still unclear (methodologically and results-wise) why do you use GP to predict the optimal control setpoints of the Helix? What is the application of it? A GP should be needed only if the output should be interpolated efficiently in many future different configurations. But in the present case of optimal setpoints selection, you actually yourself refine in the regions that are promising, so in the end your final (fixed?) choice of St $= 0.25$ is not really coming from the GP, but rather selected directly from a pool of points that you had to compute in AMR-Wind anyway. I do not see the application of the GP in Section 3 for setpoints selection, apart from drawing nice contours. In different ambient wind conditions or wind farm configuration, would the optimal setpoints change, so that you need to apply your GP to find new ones each time? But this GP was trained on this two-turbine setup, so it looks only applicable in the cases that you already ran. So there was no need for a GP, after all. I am being overly critic here to clarify my thought, but please add a general discussion on this point (why is a GP needed at all in Section 3?).

We thank the reviewer for this critical but constructive remark. The GP in Section 3 serves two complementary purposes. First, it acts as a sequential experimental design tool: at each iteration, the GP mean and uncertainty guide the selection of the next LES, focusing computational effort on the most informative regions of the parameter space. This allowed the optimal settings to be identified from approximately 20 LES runs, compared to the 50 or more runs a representative grid search over frequency and amplitude would have required. Second, the GP provides a smooth function of the power ratio surface, enabling efficient evaluation of turbine performance at unsampled conditions. Whether the optimal control setpoints change under different ambient conditions (e.g., wind speed, turbulence intensity) is not assessed in this study, and this constitutes an open research question. However, the GP framework allows one to assess, without running additional LES, whether a given combination of frequency and amplitude is expected to yield a net power gain under the current conditions and, therefore, to reason about whether deploying the helix is beneficial in a given setting. Extending the GP to include ambient conditions as additional inputs would allow this question to be addressed in future work. We have expanded the introduction of Section 3 to make these motivations explicit:

"This section presents the framework for tuning wake-mixing control settings, demonstrated for a two-turbine array simulated with AMR-Wind as described in Section 2 at a wind speed of $U_\infty = 8$ m s$^{-1}$. The GP regression framework serves two complementary purposes in this context. First, it acts as a sequential experimental design tool. At each iteration, the GP mean and uncertainty are used to guide the selection of the next LES, focusing computational effort on the most informative regions of the parameter space. This allows optimal settings to be identified from a smaller number of simulations (approximately 20 in the present study) than the 50 or more runs a representative grid search over frequency and amplitude would require. Second, the resulting GP model provides a smooth function of the power ratio surface across the full frequency–amplitude domain. This allows one to efficiently evaluate, for any given control setting, whether deploying the helix is expected to yield a net power gain, how large that gain is likely to be, and the expected variation in power gain due to the turbulent nature of the inflow. Although demonstrated here for low-turbulence conditions, the framework is applicable to other wake-mixing strategies and can be expanded to include different atmospheric conditions."

7. Are you going to share any data (simulation, trained GP, etc.) publicly with this paper, as it is promoted for Wind Energy Science? This would help to reproduce your findings or directly apply your GP models, and thus would increase the visibility of your work to the community.
   We agree with the reviewer that sharing data and code improves reproducibility and increases the visibility of the work. We will make the following available on Zenodo prior to final publication: (1) the loads database used to train the GP load surrogate model, together with the GP training and prediction scripts; and (2) the GP model for turbine power, including the averaged 10-minute power data extracted from the LES. Full OpenFAST time series and LES data are not shared due to their prohibitive size. A *code and data availability* statement has been added to the manuscript with the Zenodo reference, to be updated with the final DOI upon publication.

Minor comments:

- Eq. (7): your definition of TI is incorrect. You are missing the time-averaging operator (typically denoted as overbar, and explained what it means) over the u'u', v'v', and w'w'. Without this time-averaging, your TI is instantaneous as it is written (which does not really makes sense).
  Thank you for pointing out this mistake. Eq. (7) has been modified to include the time averaging operator. The correct definition of TI was used in the data processing, so the results remain unaffected.
- Line 110: Shouldn't the surface roughness has a unit in AMR-Wind (meters, I expect)?
  Unit for surface roughness (m) was added to Line 110.
- Line 112: Which wind speeds exactly? Would be good to specify (in a table for example). Also, what TI level, what vertical shear ? Any veer included ? I see now that (most of) this information is in Table 1, but please add already a pointer to the Table 1 here on line 112, to avoid sounding a bit vague.
  The text now points to Table 1 for an overview of all the settings:
  "An overview of all the precursor simulation settings is provided in Table 1."
- Line 114: what do you mean exactly by "generally"? In some cases, not?
  For all cases, except the 6 m/s precursor case, a simulation time of 36,000 s was used. However, the 6 m/s case required a longer time to develop due to the lower wind speed, hence the simulation time was set to 48,000 s. We changed the description accordingly:
  "The precursors were simulated between $36,000\,\mathrm{s}$ and $48,000\,\mathrm{s}$, depending on inflow velociy, to allow sufficient time for turbulence to develop and reach a quasi-steady state."

- Figure 1 caption: Please add temperature (Theta [K]) first in the caption. It's odd to list only three plots when there are four. Also, adapt the y-axis range to be uniform across all 4 plots, for direct comparison and clarity.
  Thank you for pointing out the missing entry in the caption. The caption has been changed and the y-axis is now consistent across all 4 plots.
- Lines 123-124: please be a bit more precise here where it naturally fits (OpenFAST is only a wrapper code): "the aerodynamic forces of the turbine are modeled using the AeroDyn (v15.03?) module of OpenFAST."
  This was indeed the version we used. The text has been adjusted as suggested.
- Line 126 "59 actuator points": Again here please be a bit more precise shortly. Where are those actuator points located? Are those on the blades? Is it 59/3 points per blades (which does not match a integer number)? Clarify with an additional (half-)sentence.
  We changed the sentence to clarify the matter:
  "The ALM model uses $59$ actuator points per blade, corresponding to the number of airfoil files for the IEA 22MW turbine used by AeroDyn."
- Line 132: typo OpenFAST (styled capitalized).
  This typo has been fixed.
- Section 2.2 Wind farm flow control methods: While I agree on keeping this section short and clear to the point as it is, references should be added here and there to point to published literature on both wake steering and the Helix method, to guide the interested reader to go read other solid sources on them if more details are needed.
  Secion 2.2 now contains several references to both wake steering and the helix method.
- Figure 2: Expand this caption with more information on the main elements of the figure. The reader should not have to refer to the main text to understand your figure. Especially, what is the 4Dx10D subdomain here ? Is it the whole simulation domain, or a different mesh refinement level ? Unclear here from the figure + caption. I see (later) that it's explained in the text, but the figure + caption should be self-contained.
  We completely agree that the previous caption did not properly describe the contents of the figure. The caption has been changed accordingly:
  "Schematic of a simulation setup in AMR-Wind. The simulation domain has dimensions 15.7 $D$ x 15.7 $D$ x 4.5 $D$. A single mesh refinement (4 $D$ x 10 $D$ x 2.1 $D$) is indicated by the rectangle enclosing the turbines. The two turbines are separated by a distance of 5 $D$."
- Table 1: Shouldn't the surface roughness has a unit in AMR-Wind (meters, I expect)?
  The correct unit has been added.

- Lines 142-145: This part should be inverted, to respect the physical order from causes to consequences. From a control perspective, sinusoidal individual pitch references signals are imposed on the turbine, which results in desired varying tilt and yaw moments. The way that you wrote, it sounds like the moments are the origin cause and the pitch signals are the consequence (which is physically wrong, in my understanding).
  We have changed the order and rephrased parts of these sentences to get the right order of cause and effect:
  "With the helix control strategy, individual blade pitch control is used to exert time-varying blade moments on the incoming flow. These moments are designed in such a way that, when transformed from a rotating to a non-rotating reference frame (Bir, 2008), slowly varying yaw and tilt moments are applied to the flow. These moments create the characteristic helical shape of the wake."
- Line 145 "the frequency of the pitch signals": You forget to define that it's $f_{Beta}$ here. Equation 1 does not define what is $f_{Beta}$, it should be defined in the text.
  $f_\beta$ is now introduced in the text before the equation.
- Line 145 "Strouhal number": Again, define here that it's denoted $St$. Equations (2) is unclear otherwise.
  $St$ is now introduced in the text before the equation.
- Line 150 "excitation frequency": just above you name it actuation frequency, and now here excitation frequency. Please unify for clarity. I think "excitation frequency" is the most correct in both here, so please correct on line 149 before the equation. Actuation should be more the used term for the $f_{Beta}$.
  Both actuation frequency ($f_\beta$) and excitation frequency ($f_e$) have now been defined in the text.
- Line 150-151: This sentence is unclear, I needed to read it five times. It would be much more straightforward and clear to say: "In case of the counterclockwise helix implementation, the excitation frequency is added to the rotational frequency, while in the clockwise version it is subtracted from it."
  To avoid any unnecessary confusion, we have changed this sentence as the reviewer suggested.

- Figure 3: Here you should more rigorously define the symbols used. What are Theta 1, 2, 3? You haven't defined that it's the individual pitch angle of each blade. Furthermore, here you use a Helix amplitude $A_{Beta}$ of 2° (or 4°?) as example. Later you talk about the cases of $A_{Beta}$ that you considered, but here in this part is missing an equation showing how $A_{Beta}$ is integrated into the pitch signal. All of this might be obvious to you, but not to a reader non-initiated to the Helix method.
  To clarify how the helix was implemented by the pitch controller, we have added Equation (3) for the pitch angles of all three blades as a function of time. Furthermore, in Figure 3, we now indicate both the pitch amplitude $A_\beta$ and the actuation period $T_\beta = f_\beta^{-1}$.
- Line 182 "one-second interval": so your wind fields for the surrogate training have a time step of 1s. Is that accurate enough to capture the high-frequency effects on DELs ? (I am doubting). Please comment on this point. After simulating your wind farm in LES at 0.05s timestep (which probably was expensive), wasn't it easily doable to extract the wind slices at higher frequency than 1Hz?
  The reviewer raises a valid point. To assess the impact of the 1 Hz inflow sampling frequency, we compared the Power Spectral Density (PSD) of two load channels between the AMR-Wind+OpenFAST simulation (20 Hz inflow) and the standalone OpenFAST simulation (1 Hz inflow), shown in Figure 1 of this document. For the tower base moment (TwrBsMyt), the two simulations agree closely across the full frequency range. For the blade root flapwise moment (RootMflp), agreement is good up to approximately 15 Hz, above which the LES inflow excites structural content absent in the standalone simulation. However, the PSD magnitude in this band is approximately 70–100 dB lower than the dominant low-frequency content, and its contribution to the total DEL is expected to be small. The 1 Hz sampling frequency was primarily chosen due to data storage constraints, each flow plane is a time-series of more than 3 GB when extracted from AMR-Wind. Resolving the higher frequency load components in the OpenFAST simulations would have required at least 10 times the amount of data, which was not feasible in our case. We have added a sentence in Section 2.4 acknowledging this limitation:
  "It should be noted that the 1-Hz sampling frequency of the extracted flow slices could limit the high-frequency inflow-driven load content, and some structural excitation present in a full LES-driven simulation may not be captured by standalone OpenFAST simulations at higher frequencies. However, sampling the inflow at a frequency similar to that of the LES was not feasible from a data storage perspective."

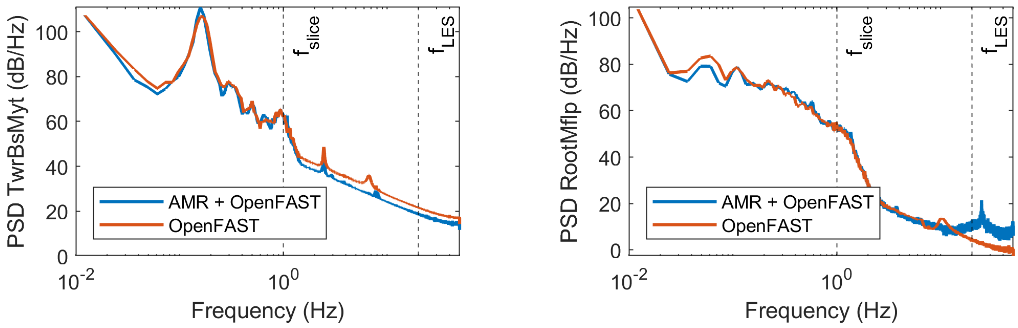


Figure 1: PSD of the tower base and blade root flapwise bending moments. The plot compares the load signals of the AMR-Wind + OpenFAST (20Hz inflow) simulations and the standalone OpenFAST (1 Hz inflow) simulations.

- Line 193-194: you haven't properly defined where/how the amplitude $A_{Beta}$ influences your control actuation (for readers non-initiated to the Helix).
  Following the definition of the pitch angles in Equation (3), we have added the following sentence to explain how the amplitude affects the turbine and the wake: "Higher pitch amplitudes result in larger wake deflections, but they are accompanied by higher power losses and increased structural loading (Taschner et al., 2023)."

- Figure 4: "Schematic of the load surrogate model": It should rather be "Schematic of the process to train and validate the load surrogate model used in this study". Since the load surrogate model per say is only a subblock in Figure 4.
  Caption has been changed as suggested by the reviewer.

- Equation 6: what is u(Theta, r) exactly ? You should first define that it's the local time-averaged wind speeds for the streamwise direction. The time averaging is done before the space averaging of Equation 6, I presume.
  Thank you for pointing this out, this is indeed the time-averaged wind speed. We have modified the equation to include the time-averaged wind speed $\bar{u}$.

- Line 207 "This was done to distinguish between the wake of a turbine operated with greedy control and the helix method, as the latter generally causes higher velocity fluctuations.": This explanation is a bit unclear and not really justifying the choice of $U_{\mathrm{inf}}$. In the wake of a helix-actuated turbine, there is higher velocity fluctuations but also higher mean wind speeds (if the Helix is effective to enhance wake recovery). So if you chose to normalize TI by the local wind speeds instead of $U_{\mathrm{inf}}$, the values of TI would have been more balanced. I am not saying to change your choice, but clarify the discussion here. If you start an explanation of your choice (which I appreciate), make it complete.
  As the reviewer explains, in the wake of a helix-actuated turbine, we expect higher mean wind speeds, but also larger velocity fluctuations. If we would use the local average wind speed to determine the TI, we observe very similar TI values for the wakes of greedy-controlled turbines and helix-actuated turbines. This would make it impossible for the load surrogate model to predict the increased loading of a turbine operating in a helix wake. We have revised the previous explanation to make it clear that the load surrogate model requires higher TI values to predict loads in a helix-actuated turbine wake.
  "In this case, the TI sectors provided as input to the load surrogate model are higher for turbines operating in a helix wake. Hence, the surrogate model can predict the additional fatigue loading resulting from the helix wake."
- Table 3: you list the rotor speed as an output, but there isn't any results of rotor speed presented in this paper. I haven't checked for each output channels you list, but I consider that only results that are included in the paper should be listed here (the matter is less of output channels from your simulations, that actual selected channels that you show in the paper, I believe). Furthermore (caught my eye), why do you use the unit s$^{-1}$ for rotor speed instead of the most commonly used RPM (in wind energy literature and in OpenFAST)? I would just delete this line (and maybe other) from Table 3 if it's not used in the rest of the paper.
  As per the suggestion of the reviewer, we have removed some of the OpenFAST channels from Table 3 that are not used or discussed in the results section of the paper.
- Line 214 "$N_{bins} = 200$": you use the subscript bins here but not inside Eq. (8), please unify. Here and everywhere in the paper, check that all symbols are cleanly and unambiguously defined before or where they are used.
  Thank you for pointing out this omission in Equation (8), it has been corrected.
- Line 215: "$N_{cycles}$ refers to the number of 1Hz cycles in the 10-minute time series": Add here ", as per convention,". This choice of using "1Hz DELs" is a common convention in our domain, but other choices could be valid (some literature use for example 2e6 cycles over 20 years operation, which just scales differently the DELs).
  The reviewer's suggestion has been added to the sentence.

- Line 255: "preprocessing": shouldn't it be rather "post-processing"?
  Adjusted to 'post-processing'.
- Line 256 "Matern covariance function": add a reference for it.
  Reference to Rasmussen and Williams (2006) was added.
- Line 273 "The next step in the modeling framework is to expand the GP model to include pitch amplitude.": Why did you do it iteratively? Wouldn't it be more efficient and more accurate for the GP to directly learn on both the Strouhal and the pitch amplitude? Are there good reasons to decouple them as you did?
  The reason for this approach is twofold. First, the initial Gaussian Process model with only Strouhal number as input allows for better demonstration of the model, by showing how it is updated after each simulation. Second, the initialization of the model requires less data in the case of a single input. However, when we proceed to the GP model with two inputs, the Strouhal number and amplitude are not decoupled. All prior data is being used to train the new model. We have modified parts of the text to make this point clear:
  "For the sake of demonstration, we first focus on a single input GP model where the Strouhal number is the free variable, and the pitch amplitude is fixed at a value of $A_\beta = 2.5°$. Using only one input also allows for easier initialization of the model, when the amount of data is still small."
- Eq. (12) is just a repetition of Eq. (1) and (2). Please just point to the first ones. Also because it's odd to define equations in your results section 3., it should all be contained in methods section 2.
  We agree with the reviewer that this equation is redundant and have removed it from this part of the paper.
- Line 315 "averaged over the six turbulence seeds to obtain a steady-state representation of the fatigue loads.": that goal is unfeasible, per the very definition of fatigue loads (that are time-varying). What you mean is: "averaged over the six turbulence seeds to obtain converged quantities of the fatigue loads, helping towards predicting them from a steady-state representation of the inflow."
  Thank you for this suggestion, which has now been adopted in the paper.

- Line 385 "However, the average trend is similar to that of T1 for most components.": I think this statement is a bit misleading, since the root causes for T1 and T2 are so different (respectively increased actuations, and more turbulent wake). Furthermore, the statement is not correct for the tower base fore-aft DEL according to your results Fig. 12, where it increases for T1 but decreases for T2.
  We have adjusted this part to discern the cause of the additional loading of both turbines:
  "For increasing pitch amplitudes, the average power gain increases, but so do the loads of each component due to the periodic pitch actuation. The results from the downstream turbine show greater load variations across changing inflow conditions. Similar to T1, we also observe higher loading for increasing pitch amplitudes, due to the additional turbulence introduced in the wake. Only in the tower base load channel we see a decrease in loading as pitch amplitude increases, indicating that not all loads are driven by turbulence but can also be affected by the average inflow wind speed."
- Line 393-395 "With AMR-Wind, a pitch...": this whole last sentence raises an important point without clarifying it enough. Isn't AMR-Wind using OpenFAST to simulate the turbine aerodynamic and structural behavior? For an upstream turbine, how can the results differ so much? And how can you know which one to trust more? You write "the power loss of T1 is overestimated" but that clearly depends on which result you consider more accurate. Please add clarifications here.
  The two simulation approaches differ fundamentally in how blade aerodynamic loading is computed: the AMR-Wind–coupled simulations derive blade forces directly from the LES-resolved flow field via the actuator line method, whereas the standalone OpenFAST simulations rely on blade-element momentum theory with a dynamic inflow correction. This modeling difference is the primary reason the two approaches yield different magnitudes of power reduction under dynamic pitch actuation. Each approach involves its own modeling assumptions and associated uncertainties, and we do not consider either framework to provide an unambiguous ground truth for this comparison. We have revised the manuscript to present the two results as reflecting different modeling choices rather than one being more accurate than the other:
  "With AMR-Wind, a pitch amplitude of $4^\circ$ results in a power loss of 2%, whereas the same pitch amplitude results in a power loss of 5% when simulated with standalone OpenFAST. This difference is attributed to the different aerodynamic modeling approaches underlying each simulation environment. Where AMR-Wind resolves blade loading directly from the LES flow field via the actuator line method, standalone OpenFAST relies on blade-element momentum theory with a dynamic inflow correction. As the surrogate model for generator power is trained on standalone OpenFAST data, this modeling difference propagates directly into its power loss predictions for T1."

# Response to comments of Reviewer #2

I found the paper interesting and enjoyable to read. It presents a valuable advancement of the helix approach for wind farm control, particularly through the use of GP-based surrogate modelling and the combined consideration of power extraction and structural loading. I especially appreciated the section in which energy yield and loading are combined using Gaussian Processes, as this provides a useful perspective on practical wind farm control. Overall, I believe the manuscript is suitable for publication in Wind Energy Science after satisfactory minor revisions. My comments are provided below, divided into major and minor points.

1. From reading the paper, it is not particularly clear to me why wake mixing is a promising approach. It is quite complex to simulate, requiring higher-fidelity simulations, it involves a dynamic actuation, which introduces additional challenges for simplified modelling compared with other control approaches, and it systematically increases loading. I am not saying it is not promising, just that it is not particularly clear from reading the paper.
   We have expanded the introduction to more clearly motivate the potential benefits of wake-mixing control relative to alternative WFFC strategies. Wake mixing has been shown to increase power performance [2, 4], and under low wind veer conditions, has even been found to outperform wake steering [3]. Furthermore, a recent study showed the helix method to be more robust to wind direction variations than wake steering in the case of full turbine alignment [1]. We also clarify that the associated increase in complexity and fatigue loading is precisely why an efficient, load-aware modeling approach, which is one of the contributions of this paper, is needed to evaluate whether the trade-off is worthwhile in a given scenario.
2. Section 2.3: I find this section very specific and not very explanatory for the common reader. I believe it could be helpful to introduce at the beginning of the section a higher-level explanation of Gaussian Process regression, mentioning in a clearer and simpler way the aims, the steps (initial dataset generation, training, then predictions), and the applications. Section 2.4: I find this section particularly difficult to follow. I suggest first adding a short summary for the reader to understand the general concept. Then, I suggest describing the various steps in order following the workflow shown in Figure 4 in a simple and systematic manner. As an example, I find the description of the same method significantly clearer in the Conclusions.

Thank you for this suggestion. We have added a higher-level introduction paragraph at the start of Sect. 2.3, explaining the aim of GP regression in this framework (constructing a cheap surrogate for a relationship that would otherwise require expensive simulations), the different steps required to obtain this model (generating training data, training the model, predicting at new inputs), and the two applications of this framework in the paper (controller tuning in Sect. 3, and load prediction in Sect. 4). Similarly, we have added a short summary at the start of Sect. 2.4 that walks through the load surrogate workflow shown in Fig. 4 chronological order, before the detailed description of each step.

3. I think there is a substantial difference between employing GPs to represent a dimensional space as best as possible for interpolation (which appears to be the aim for the loading part of the paper) compared to using GPs in the context of optimisation, where the aim is to finding the optimum of an objective function (aim for the controller tuning). In the latter approach, accurate predictions across the entire design space are not required, but it is more about balancing exploitation and exploration. Please clarify this important point in the manuscript. In lines 400 and 401, "First, the power output of a two-turbine array was modeled as a function of the actuation frequency and amplitude." mentions modelling of the function, so what is the aim? Modelling, optimisation, or both? Please clarify.
   We agree with the reviewer that this is an important distinction. We have clarified in the manuscript (Sect. 2.3) that the two applications of the GP framework place different demands on the model. For the load surrogate model, the training data are generated in advance from a fixed database, and the aim is to obtain an accurate representation of fatigue loads across the full range of inflow and control conditions, since predictions may later be required for arbitrary combinations of these conditions. For the controller tuning framework, the aim is twofold. First, we want to obtain a model that shows where and to what extent the helix method is effective across the frequency–amplitude domain, and second, use this model to make informed decisions on which settings to simulate next. The following text was added to Sect. 2.3:
   "These two applications place different demands on the GP model. For the load surrogate model, the training data are generated in advance from a fixed database of simulations, and the aim is to obtain an accurate representation of the fatigue loads across the full range of inflow and control conditions, since predictions may later be required for arbitrary combinations of these conditions. In the controller tuning framework, the GP serves a dual aim, which is to obtain a model that shows where, and to what extent, the helix method is effective across the frequency–amplitude domain, and to use this model to make informed decisions on which settings to simulate next."

Furthermore, we extended the original Line 400 to make the aim more explicit: "First, the power output of a two-turbine array was modeled as a function of the actuation frequency and amplitude, using the GP both to characterize where and to what extent the helix method is effective and to guide the selection of simulation settings toward near-optimal control settings."

4. Lines 261-262: "Based on the inferred power ratio, a new frequency is selected and subsequently applied in a new LES run, after which the process is repeated.". Could you please further explain and justify the criteria used to select new frequencies to simulate? As mentioned in the point above, I am confused if the aim of this power production section is modelling or optimisation. From my understanding, determining the location of the optimum control inputs, namely (i) fixed amplitude with variable frequency and (ii) combined frequency and amplitude variations, to maximise power production of the overall farm is at least one of the aims. My doubt is why an optimisation algorithm is not employed, such as gradient-based optimisers, genetic algorithms, or Bayesian optimisation. The latter in particular aims at balancing exploitation and exploration through GPs, effectively locating new data points to simulate through an acquisition function by exploiting the GP-predicted maximum and also exploring areas with high uncertainty. I believe manually locating the new locations is quite inefficient and ineffective. With the two main components of model-based optimisation being the model and the optimiser, I believe the latter is missing in this paper. Please clearly state the employed optimisation procedure and the limitations of the proposed methodology, which of course could be further expanded in future work.
   We have clarified in the manuscript (Sect. 3) the criteria used to select new simulation settings: the frequency (and later, frequency–amplitude) range was first explored broadly to characterize the shape of the power ratio function, after which subsequent simulations were concentrated in the narrower range identified as most promising. We have also added an explicit justification for not employing a formal optimization algorithm such as Bayesian optimization: given the multi-day computation time of each LES, the time required to manually inspect the GP model and select a new setting is negligible in comparison, and the one- or two-dimensional design space considered here can be visualized and interpreted directly. We agree that an automated, acquisition-based approach becomes increasingly valuable as the dimensionality of the design space grows beyond what can be visualized directly, for example when additional ambient conditions are included as GP inputs, and we now state this explicitly as a consideration for future work.

"Formal optimization algorithms, such as Bayesian optimization, provide a systematic way to balance exploration and exploitation through an acquisition function, and would be a suitable choice for this problem. However, we consider this most valuable either when the selection of new points must itself be fast and automated, or when the design space cannot be visualized directly, i.e., when more than two inputs are considered. In this study, each LES run required several days of computation time, so the time needed to manually inspect the GP model and select a new simulation setting is negligible in comparison, and the one- or two-dimensional design space considered here could be visualized and interpreted directly. We therefore consider the manual selection approach adopted here to be an appropriate choice for the present study, while acknowledging that an automated, acquisition-based approach such as Bayesian optimization would become increasingly relevant as additional inputs (e.g., ambient conditions) are added to the model and the design space can no longer be easily visualized."

5. I believe the paper lacks information regarding the applicability of the proposed methodology to wind farm control. First, the computational time required to generate the datasets and to perform the Gaussian Process hyperparameter tuning is not reported. As control applications move towards larger wind farms and, ideally, closed-loop scenarios, understanding the computational effort associated with dataset generation, surrogate-model training, updating, and querying is important for assessing the practical appeal of such approaches, both in academia and in industry. Second, I suggest providing more detail on the size of the datasets used, for example the number of training points, as is already done for the loading scenario. Finally, Gaussian Process training scales cubically with the number of training samples, which can make their application challenging for larger datasets. Since the paper discusses the possible inclusion of additional dimensions or operating conditions, such as those mentioned in line 305, I believe it would be valuable to acknowledge the scaling behaviour of GPs and the associated challenges and limitations that may arise. I am not necessarily asking the authors to provide all of this information in full detail. Rather, I suggest that the authors address these points in the manuscript, at least by acknowledging their importance and discussing their implications for the practical applicability of the proposed methodology.

Computational cost is now reported explicitly throughout the manuscript. Each LES requires one to two days of runtime, with post-processing and aeroelastic simulations after each LES case adding a further several days. Once the loads database is available, GP hyperparameter fitting takes 10–20 minutes per load channel, while the smaller controller-tuning GP (in the order of 120 training points) fits in seconds. Dataset sizes are also now reported for both applications. The controller-tuning framework resulted in 20 LES runs, totaling approximately 120 training points. Finally, we have added a discussion of the cubic scaling of GP training with dataset size, noting that the load surrogate model, which already uses ten inputs and several thousand training samples, remains computationally manageable at the scale considered here. However, we acknowledge that this scaling could become a practical constraint as the framework is extended to cover a wider range of conditions, motivating the use of sparse GP methods in future work.

6. In general, I found quite a few inconsistencies in terminology, symbols which are not introduced, and the nomenclature not is not always clear. Please address these throughout the manuscript.
   Thank you for addressing this concern. We have systematically revised all equations and results sections to ensure every symbol is defined on first use. An appendix with nomenclature and abbreviations has also been added to the paper.

7. Several figure captions are not self-contained, as the reader is required to go back to the main text for clarifications. Please address this.
   We have revised the captions throughout the manuscript to make them self-contained, so that a reader can interpret each figure or table without needing to consult the main text. This includes Figs. 1, 2, 3, 4, 6, 7, 8, 9, and 12, and Table 3.

8. I believe the use of subfigure labels, together with explicit references to them in the text, would significantly improve the readability of the paper. It can be quite challenging to follow the discussion of the results and trying to understand which subfigure it refers to. I found this to be particularly relevant for Figure 6, 7, 9, and 11.
   We have added subfigure labels, together with explicit references in the text, to Figs. 6, 7, and 11 to improve readability. For Fig. 9, the four panels are already identified in the text by their load channel names rather than by position, so a reader can directly match each panel to its corresponding discussion without needing an explicit subfigure label. We therefore consider the current presentation sufficiently clear for this figure.

9. Are the dataset and code associated with this paper publicly available? I think this would benefit the wind energy community and improve reproducibility.
   As previously mentioned in response to a similar question by reviewer 1, we will make the following available on Zenodo prior to final publication: (1) the loads database used to train the GP load surrogate model, together with the GP training and prediction scripts; and (2) the GP model for turbine power, including the averaged 10-minute power data extracted from the LES. Full OpenFAST time series and LES data will not be shared (except by reasonable request) due to their prohibitive size. A *code and data availability* statement has been added to the manuscript with the Zenodo reference, to be updated with the final DOI upon publication.

Minor comments:

- Please include a sentence in the abstract summarising the results from the loading assessment. Currently, only the method capabilities are mentioned.
  The following sentence was added to the abstract:
  "Applying the load surrogate model to a two-turbine case study demonstrates the trade-off between additional power gain and increased structural loading of both turbines with the helix method."
- Lines 19-20: "Wind farms... wakes". Please slightly elaborate, or clarify turbulent wakes arise from aerodynamic interactions between turbines, for non-expert readers.
  A brief description of wakes has been added:
  "These wakes arise as turbines extract energy from the incoming flow, leaving a zone characterized by lower wind speeds and higher turbulence."
- Line 21: I would replace "or" with "and/or" as the two objectives are not necessarily mutually exclusive.
  Changed "or" into "and/or".
- Lines 56-57: "load surrogate models offer a more efficient solution". Please clarify what it is more efficient than. LES coupled with aeroelastic simulations? Also, consider mentioning the distinction between efficiency in prediction and the requirement for a dataset generation and a surrogate model training.
  We modified Line 56 to be more explicit:
  "To assess the fatigue loading to which wind turbines are subjected under different conditions (i.e., atmospheric, operational, wake effects), load surrogate models are a more efficient alternative to running expensive simulations for each case."
  Furthermore, we added a sentence on the distinction between training and prediction as suggested by the reviewer:
  "Although these models require a large number of simulations for training, once trained, they enable efficient prediction and offer significant flexibility."

- Lines 58-59: Did Dimitrov et al. create a dataset for these site-specific inputs and then evaluate different surrogate models? In line 61, "this dataset" is mentioned but never introduced.
  "This dataset" refers to the previous sentence and is indeed the dataset created by Dimitrov et al. Line 58 was rephrased to make this clear:
  "Dimitrov et al. (2018) used various site-specific ambient conditions (e.g., wind speed, turbulence intensity, shear exponent, wind veer) to construct a load database and used these conditions as inputs to compare multiple surrogate models for predicting fatigue loads."
- Line 64: Is the LUT constructed from data generated using aeroelastic simulations? The description is generally vague and could be confusing for some readers.
  Yes, this lookup table was constructed from aeroelastic simulations. This has been added to this sentence:
  " A similar approach from Mendez Reyes et al. (2019) used wake parameters such as depth, width, and lateral spacing between the wake and the rotor to build a multidimensional lookup table (LUT) based on aeroelastic simulations."
- Line 81: Please replace "yield" with "energy yield" for clarity.
  "Yield" has been changed to "energy yield".
- Line 89: I believe it can be helpful to clarify early in the paper what variations in conditions the array will experience. Are the conditions atmospheric, operational, both?
  Thank you for this suggestion, we have added the following:
  "Evaluating WFFC performance in terms of energy yield and fatigue loads with the helix method on a two-turbine array under different operational conditions in terms of the helix pitch amplitude and array alignment."
- In Figure 1 caption, please mention the temperature profile shown in the figure.
  The caption now also mentions the temperature profile, thank you for pointing out this omission.
- Line 145: Please introduce the symbol for the frequency of the pitch signals when first mentioned.
  The symbol is now introduced when pitch frequency is first mentioned.
- Lines 145 and 148: To avoid confusion, please keep a consistent terminology. In this case, rotational frequency of the turbine is referred to as rotor frequency just a few lines after. Please check the terminology throughout the manuscript.
  "Rotational frequency" is now used consistently throughout the manuscript.
- Line 179: Load channels are not introduced here or mentioned before. Please add a short explanation.
  We agree with the reviewer that it makes sense to introduce the load channels at this point. Therefore, we have moved Table 3 before Table 2 and mention the list of load channels in Line 179.

- Line 188: How is "no wake overlap" defined considering the variations in operating and atmospheric conditions? I would imagine that already between the helix and baseline cases, wake expansion would be different. Please clarify.
  In this case, the no wake overlap condition occurs at small turbine spacing up to $5\,D$ with the baseline control case. As the reviewer notes, this condition may not always apply across different control strategies. We have clarified this in the text: "... to no wake overlap (at close turbine spacing with baseline control and for $\Delta y = \pm 390\,\mathrm{m}$ ≈1.37 $D$)".
- Line 185, 187, and 188: In the lines mentioned, please also express the grid and other measurements normalised by the rotor diameter for clarity.
  Normalized values have been added to the dimensions.
- Line 191 and 192: "The OpenFAST simulations for the loads database also include the turbine using different control strategies to assess the impact on loads.". Weren't the simulations with control actions just specified above? I might have misunderstood, please clarify as this is not clear to me.
  In this paragraph, we make the distinction between LES cases, which were used to extract the flow slices, and the standalone OpenFAST simulations that were run to generate the input for the load database. We have added some clarification to this part of the text:
  "In the LES cases used to obtain the flow slices, we employed different control strategies, including greedy control, wake steering, and the helix method. Furthermore, several simulations with three turbines, each separated by a distance of 4.5 $D$, were conducted to include the effect of accumulated turbulence in the wake. An overview of all the control strategies that were simulated is provided in Table 3. Each of these control cases was simulated for all the precursor wind speeds specified in Table 1."
- Line 260: Please define the power ratio as it has not been previously introduced or explained. Also, please be consistent with the power ratio definition in the captions of Figure 6 and 7.
  The power ratio is now defined as ($\frac{P_{\mathrm{WFFC}}}{P_{\mathrm{BL}}}$), and also included in the caption of Figs. 6 and 7.
- In Figure 6 caption, please mention what the different LES samples correspond to in terms of variations, range, and interval if applicable.
  Each LES is divided into six 10-minute intervals, which are subsequently averaged. The caption now specifies this:
  "Each sample corresponds to a 10-minute average taken from the 1-hour LES timeseries."
- Please be consistent when referring to the power ratio (for example, see the difference in y-axis between Figure 6 and Figure 7).
  Thank you for pointing this out, Fig. 6 now has the same label as Fig. 7.

- Figure 7: labels (a), (b), and (c) are mentioned in the caption but are not shown in the actual figure. (I find this plot very nice!)
  Thank you. The (a), (b) and (c) labels were included in the figures, but they were indeed difficult to discern due to the background color. We have moved the labels to a different position, so they are easily visible.
- Line 274: I believe "hyperparameter" is incorrect here as the inclusion of the pitch amplitude corresponds to an additional GP input dimension (or input variable). Please correct this accordingly.
  The reviewer is correct that the GP is extended with an additional input dimension, but this also requires an additional input lengthscale (hyperparameter) to be optimized. The sentence has been changed to clear this up:
  "‥, consisting of adding an additional input ($A_\beta$) and a corresponding input lengthscale (hyperparameter)."
- Lines 279-281: In "The optimum indicates that at some point, the power loss of the actuated turbine (T1) can no longer be compensated by the power increase of the downstream turbine(T2).", shouldn't it be the opposite? The power gain from T2 exceeds (at its maximum) the power loss experienced by T1.
  In Fig. 7, we see that the optimum for the two-turbine array is different from the optimum observed for T2. So yes, the power gain from T2 still exceeds the loss of T1 for pitch amplitudes above $4^\circ$, but the net gain is smaller than in the case of the optimum settings. We have slightly rephrased this sentence to clarify:
  "The optimum indicates that for pitch amplitudes above $4^\circ$, the additional power loss (with respect to the optimal settings) of the actuated turbine (T1) cannot be compensated by the increased power of the downstream turbine (T2)."
- In Figure 7, the uncertainty of the GP optimal prediction is quite high (ranging from approximately 5% to 10% for the overall power ratio). This is quite a large difference in prediction when quantifying the capabilities of a wind farm control strategy. As mentioned in one of the major comments, the use of an optimiser could reduce this uncertainty effectively. Please clearly state the limitation of the optimum prediction in the manuscript.
  The uncertainty shown in Figs. 6 and 7 follows directly from the LES samples. For a given setting in Fig. 6, for example $St = 0.3$, we see samples (10-min average data) ranging from less than 1% to around 7%. All the samples for this setpoint were taken from the same simulation, but at different time instances. This merely indicates that, depending on the time-varying inflow conditions (different turbulent structures), the effectiveness of the Helix method can vary over 10-minute periods. Additional simulations at the same setpoint could reduce uncertainty, but this is not guaranteed with an optimizer.
- Figure 12: Please define the relative power increase in the caption, rather than only in the main text.
  Definition of relative power increase ($[P_{\text{WFFC}}/P_{\text{BL}} - 1] * 100\%$) was added to the caption.

- Line 367: please add more details regarding the simulation environment used to produce the “generator power” dataset. In line 395, it is specified that it is different from AME-Wind, so please explain here when first introduced.
  Similar to the load channels, we also trained a surrogate model for predicting power using the power measurements from the standalone OpenFAST simulations. This is now mentioned in the line referred to by the reviewer:
  “As with the various load channels, we also trained a surrogate model to predict generator power (extracted from the OpenFAST simulation database) based on inflow sectors.”
- Wake steering is included in the tuning data, but I believe it is not presented or discussed in the results. What is the aim of running these additional simulations? I might have missed it, in which case I apologise. Please clarify.
  We agree that the paper did not sufficiently motivate the inclusion of wake steering simulations. This was done for future applications, where multiple control strategies can be selected. A sentence has been added to the final part of the conclusion that mentions this:
  “Moreover, since the load surrogate model was also trained to account for yaw misalignment and skewed inflow, it offers flexibility for application to other WFFC strategies beyond the helix, such as wake steering, which was not directly assessed in this study.

# References

[1] M. Baricchio, D. van der Hoek, T. Dammann, P. M. O. Gebraad, J. Iori, and J.-W. van Wingerden. Combining wake steering and active wake mixing on a large-scale wind farm. *Wind Energy Science Discussions*, 2025. in review.

[2] J. A. Frederik, B. M. Doekemeijer, S. P. Mulders, and J. W. van Wingerden. The helix approach: using dynamic individual pitch control to enhance wake mixing in wind farms. *Wind Energy*, (May), 2020.

[3] J. A. Frederik, E. Simley, K. A. Brown, G. R. Yalla, L. C. Cheung, and P. A. Fleming. Comparison of wind farm control strategies under realistic offshore wind conditions: turbine quantities of interest. *Wind Energy Science*, 10(4):755–777, 2025.

[4] E. Taschner, A. A. W. van Vondelen, R. Verzijlbergh, and J. W. van Wingerden. On the performance of the helix wind farm control approach in the conventionally neutral atmospheric boundary layer. *Journal of Physics: Conference Series*, 2505:12006, 2023.