Answers to the comments of the reviewer 1 for the manuscript *Load Estimation in Onshore Wind Farms Using Surrogate Modeling and Generic Turbine Models* submitted to Wind Energy Science

Dear Prof. Vidal,

The authors would like to thank the reviewer for the detailed comments and suggestions to improve the manuscript. In the following, the comments are provided in italic letters. Each point is answered in non-italic letter. Changes to the text of the manuscript are reported in red.

With kind regards,

Alexander Mönnig, Ansgar Hahn, Astrid Lampert and Ulrich Römer

Reviewer 1

General Comments

Comment:
This paper investigates the use of generic turbine models combined with surrogate models to predict loads on real-world turbines, validated through comparison with multi-year measurements from an operational wind farm. The motivation of the paper is highly relevant, addressing a topic of growing interest in the wind energy community. The manuscript is overall well structured and generally well written.
Answer: We would like to thank the reviewer for this positive judgment.

Comment:
However, the current performance of the surrogate models does not appear sufficient to draw robust conclusions about the method's applicability. I recommend the authors either reframe the manuscript as a methodological paper or extend the work on the surrogate models to improve their predictive capabilities. Specific suggestions are provided below.
Answer: To underline the scope of the manuscript we have modified formula-

tions in the abstract and the introduction, to emphasize that the paper primarily investigates the feasibility of using surrogates in connection with generic wind turbine models.

General Writing

Comment:
Throughout the paper, please integrate references more smoothly into the text rather than relying on "see" or "cf.".

Answer: We appreciate this comment and have revised the manuscript with more natural formulations.

Abstract

Comment:
It is a great short and direct abstract. However, it would benefit from more emphasis on the key challenge the authors are trying to tackle here: the lack of access to commercial turbine models and whether generic models can address high-level design considerations.

Answer: Thank you for pointing this out. In the revision we have emphasized this aspect:

This article investigates the feasibility of developing surrogate models, based on slightly adapted generic turbine models, for predicting loads on real-world wind turbines. The approach aims to reduce reliance on proprietary turbine models, typically unavailable to industry practitioners, and is intended to support high-level assessments and decision-making.

Introduction

Comment:
The introduction is overall good. For clarity, please address the limited access to accurate turbine models first (before the surrogate modelling), as this is the core novelty of the work.

Answer: We agree with this and have implemented this change.

Comment:
Please include a brief literature review on the use of generic turbine models in other studies.

Answer: As suggested we added a short literature review.

Section 2: Surrogate Models

Comment:
Line 100: Please elaborate on the methodological extension related to the pitch angle.

Answer: We agree that a more detailed explanation would be helpful at this point. The new part reads:

For above-rated conditions, where a change in wind speed no longer results in a change of power, the required derivative can no longer be computed. We therefore chose to adapt the method to use the pitch angle β instead, which in pitch regulated machines continuously increases once rated power is reached. Therefore, it represents a suitable proxy for estimating the wind speeds standard deviation as

$$V_{\mathrm{std}} = \frac{\beta_{\mathrm{std}}}{\left(\frac{d\beta}{dV}\right)_v \cdot B}. \tag{1}$$

Comment:
Line 121: The reference to Dimitrov et al. (2018) is abrupt. Adding more context would make it smoother.

Answer: Following the reviewers recommendation we have reformulated the text as:

For the wind speed, a Beta distribution could be employed, as already suggested by Dimitrov et al. (2018) who observed that several of their surrogate input variables that depended on the wind speed showed higher variability at lower wind speeds. More samples are therefore needed in that region, making the beta distribution a suitable choice.

Comment:
Line 124: the reference to Mara and Becker should also be introduced with more context.

Answer: In this case, the reference to Mara and Becker is included only because their paper provides the formula we apply; we therefore did not expand the context further, but reformulated slightly.

Comment:
Line 139: The introduction of the forum is somewhat weird. Maybe the best way to do it is to remove this reference and explain this as the procedure you will follow with some justifications. It might also improve clarity to reorganize this as a dedicated section titled "Adapting the Generic Turbine Models," providing a more detailed explanation of each stage of the parameters updates. This section could be just after the introduction.

Answer: Thank you for pointing this out. The reference as been removed and the procedure is now described in more detail, having added some explanations for the required steps. However, the position of the explanation has been kept to reflect its order in the workflow/method.

Comment:
Line 150: Please clarify what error thresholds or maximum acceptable errors define the model applicability.

Answer: Previous studies of aeroelastic simulations with detailed turbine models report mean errors for flapwise bending moments and tower top tilt moments typically in the range of 10-20 %, though in some cases the errors can reach 40 %, depending on the quality of the inflow definition. With this in mind, it appears reasonable to expect the errors with additional structural and aerodynamic uncertainties, at least in the range of 50 %. We would argue that these could be still considered useful for initial load estimations and trend analyses, even though they would clearly not be sufficient for detailed design or certification purposes.[1]

Comment:
Line 152: Mentioning SPCE as a future research direction is a bit early in the paper. Consider moving this to the discussion or conclusion, elaborating on its potential. References to other approaches should either be briefly presented with names and results at some point in this paper or not mentioned at all.

Answer: Thank you for the comment, we have remove the phrase and added More advanced surrogate modeling approaches could be pursued. For example, a follow-up study tested stochastic polynomial chaos expansion (SPCE), but it did not provide significant improvements over the approaches presented here. to the conclusion.

Section 3: Case Study

Comment:
Line 205: Please clarify the limits under which generic turbine models can be considered applicable to real-world scenarios.

Answer: Thank you for your comment. This point has been addressed in the discussion above.

Comment:
Line 224: citations to be improved.

[1] M, M. Pedersen, T. J. Larsen, H. A. Madsen, and G. C. Larsen, “More accurate aeroelastic wind-turbine load simulations using detailed inflow information,” Wind Energy Science 4, no. 2 (2019): 303–323, https://doi.org/10.5194/wes-4-303-2019

Answer: The citation has been revised.

Comment:
Section 3.2: The validation of the generic model is lacking. Before introducing surrogate models, a comparison between generic turbine loads and measurements under different DLCs is essential. If this has been done, please include at least key figures, if not, it represents a crucial missing step.

Answer: We included a brief validation of model accuracy in the appendix. However, a detailed validation of the generic turbine model against measurements under different DLCs was not feasible within this work, but is an important step for a future study. The study was intended as a proof of concept to evaluate whether the investigated method is useful and promising.

Comment:
Figure 2: This figure is currently not cited in the paper. It would improve the methodology section to reference it when discussing the joint distribution and bounds definition.

Answer: We thank the reviewer for pointing this out. A reference to Figure has been added in the Section 3.4.

Comment:
Line 258: Indicate the OpenFAST version used

Answer: Thank you for pointing this out. We now indicate the version of all the tools mentioned in the manuscript.

Comment:
Line 264: Justify the use of 30 seconds as transient time with relevant references.

Answer: As suggested we now include a reference to existing literature. The revised text reads: An additional 30 seconds at the beginning are included to account for any initial transient effects, which are subsequently excluded from the analysis, consistent with other studies such as the one by Castorrini et al. (2023).

Comment:
Line 277: Explain the motivation for using the Rosenblatt transformation with GPR. Have you tuned the hyperparameters (using grid search, random search, or Bayesian optimization)? This is expected to improve model performance.

Answer: The Rosenblatt transformation could be omitted within the GPR approach, at the same time the chosen kernel better fits uncorrelated and rescaled data. The hyperparameters have been tuned via maximum likelihood, as automatically provided by Scikit-learn.

Section 4: Results

Comment:
Line 290: The opening sentence is unclear. The sensitivity analysis on polynomial order and training set size you mentioned should be first introduced and then presented, including the maximum training set size used.

Answer: In Section 3.6 the training sets of the convergence analysis are now introduced. The other parameter varied in the study, were already included in the submitted manuscript. Furthermore, we added: Since detailed convergence behavior is not the focus of this study, we only summarize the key findings here; full figures and results can be found in Appendix B. to the introduction of Section 4.1 to emphasize that the training results are not discussed in detail, as they revealed nothing new or remarkable.

Comment:
Line 300: The current surrogate models seem not sufficiently capture dynamic and nonlinear effects, you should here refer to the other models you tried, and the database should be extended, 300 samples is very little for 4 input parameters.

Answer: This study was intended to be a proof of our concept and the sample size was kept purposefully rather small, due to budget and time constraints. Future investigations will include more samples/more replications to better capture the seed-to-seed uncertainty.

Comment:
Line 304: Clarify what the "respective other model" refers to.

Answer: We agree that this needed to be rephrased for clarity. It now reads: Furthermore, we found that the models trained to predict the mean and standard deviation of the 10-minute maxima performed notably worse than their counterparts for the 10-minute mean or standard deviation.

Comment:
Line 307: Specify which convergence analysis is being referenced.

Answer: Hopefully, it is clear now with the change of Section 3.6, where the setting of the convergence analysis is introduced.

Comment:
Line 310: Provide the mentioned error curves in the manuscript.

Answer: We integrated a convergence analysis for our existing training dataset in the appendix - including the error curves. An extension of the data was not possible unfortunately due to the scope of this study.

Comment:
Figure 3: The figure is difficult to interpret. Consider restructuring it, good examples could be Figure 3 and 4 in https://doi.org/10.5194/wes-9-1885-2024, showing training sample positions relative to measurements, followed by load predictions relative to inputs.

Answer: We agree with this comment and have split the figure into two separate ones. In addition, we have slightly rephrased a related sentence in the description to improve readability. It now reads: The plot over the turbulence intensity shows that measurements at the lower end of the tower acceleration range are not accompanied by nearby predictions, while at higher turbulence intensities several outliers occur, indicating that the surrogate struggles to capture these conditions accurately.

Comment:
In addition to that, it can be seen on the figure that many training points do not align with measurements, potentially explaining low model performance.

Answer: We hope that separating the plots improves clarity. When looking at the training data, it indeed appears that many samples lie outside the turbine's typical operating region, which may explain part of the observed deviations. However, when comparing the surrogate predictions with the measurements, the agreement is much closer, indicating that the models capture the turbine's behavior well despite these differences. However, this also indicates that the sampling procedure could potentially be made more efficient, as some of the sampled conditions in the training dataset occur only rarely—either because we are focusing on this specific turbine and/or because only a very small subset of SCADA samples is shown here. This aspect would, however, require closer investigation.

Comment:
In Section 4.1, the surrogate models predict mean values adequately (except for maximum tower-top acceleration), but their variance predictions are off. Before moving on to further analysis, the following points need to be studied:

Answer: We appreciate this overall remark and address the reviewer's points individually below.

Comment:
Quantify the error between generic model outputs under measured environmental conditions and actual measurements.

Answer: Regarding the assessment of the model error of the generic turbine model, we have now added additional data to the Appendix.

Comment:
Increase the training set size, 300 samples with four input variables is small. Please show the convergence analysis mentioned earlier and consider extending to 1000 samples for instance.
Answer: We have added the convergence analysis to our appendix. However, while we agree that 1000 samples could support our study, it was not feasible to extend the dataset in the scope of this study.

Comment:
Perform a convergence study on the number of seeds to take into account to capture the seed-to-seed uncertainty.
Answer: We agree that a convergence study on the number of seeds would provide further insight into capturing seed-to-seed uncertainty. Since a convergence study on the number of seeds is beyond the scope of this work, we now list it as a direction for future research in the conclusions.

Comment:
Discuss other surrogate models, such as neural networks, which have shown promising results in similar contexts.
Answer: This relates back to our limitations in the sample size for this initial feasibility assessment. We believe that there are not enough samples for a NN to be trained, which is why we chose GPR and PCE.

Comment:
Address the poor performance at low wind speeds, and elaborate on way to address it, expanding the dataset could be a way.
Answer: We are unfortunately not entirely clear on the focus of the question and would kindly ask the reviewer to provide us with some guidance.

Comment:
Tune the hyperparameters of the models using grid search or random search for instance.
Answer: For the PCE models, the relevant parameters are the polynomial order and regression algorithm, both of which we varied. For the GPR models, scikit-learn tunes the kernel hyperparameters automatically via marginal likelihood optimization, which should serve a similar purpose to something like a grid search. What could have been done for GPR is the test of multiple kernels. But RBF should gave use a decent default behavior, deemed to be good enough for this feasibility assessment. Overall, the focus of the work was not on generating

surrogate models and therefore, we preferred to use well-established standard methods for surrogate modeling.

Section 5: Discussion

Comment:
Please revise conclusions to reflect that surrogate models captured mean values adequately but struggled with variance and extreme events.

Answer: We agree that the conclusions needed better reflect the results of our study.

The updated text is now: Aligning with existing literature, both applied methods, PCE and GPR, were effective in capturing the mean responses. However, predictions of the standard deviations were consistently less accurate than for the means, and this issue was particularly pronounced for the targets involving maximum values. In the case of tower acceleration maxima, no satisfactory surrogate model could be obtained. In general, the PCE approach was found to be more suitable for modeling the blade loads, while GPR was more suitable for predicting the tower acceleration.

Comment:
Line 376: The expectation that increasing the number of turbulent seeds will resolve discrepancies lacks evidence. Please conduct and present a convergence analysis on maximum tower acceleration with respect to turbulent seed number and training set size.

Answer: In the scope of this study we were not able to implement such an analysis. However, to acknowledge this valid point we aimed to provide a more transparent take on the observed discrepancies.

The modified passage now states: Concerning the developed surrogate models, it can be noted that both the PCE and GPR proved to be effective and efficient at capturing turbine responses of the training data. The challenges in fitting the standard deviation models may be related to the noisy response surfaces. This, however, needs to be investigated in future studies, for instance through additional turbulent realizations and dedicated convergence analyses. Given the lower than anticipated costs for generating the current set of training data, such extensions would be realistic. Therefore, the development of surrogates as a predictive tool remains feasible for industry practitioners, and the current models provide a good baseline for more complex models that may be tested in the future.

Summary and Recommendations

Comment:
The scope of work is highly relevant, and the preliminary findings on turbine aging and inter-turbine variability are promising.

Answer: Again we acknowledge that the reviewer sees the potential in the data and analyses.

Comment:
However, substantial improvements are required before publication. Key recommendations are:

Answer: We appreciate the reviewer's overall summary and address the individual points below.

Comment:
Provide an error estimation from discrepancies between the generic OpenFAST model and measurements.

Answer: We have integrated the results of our initial model validation approach, in which some selected SCADA points where compared to OpenFAST model results.

Comment:
See the impact of increasing the size of the training dataset on the model performances.

Answer: We integrated a convergence analysis for our existing training dataset in the appendix. An extension of the data set was not possible due to the scope of this study.

Comment:
Elaborate on surrogate model hyperparameter tuning, considering grid search or random search.

Answer: This point has been addressed during the discussion above.

Comment:
Expand the discussion on alternative models tested, consider implementing neural networks with larger datasets.

Answer: We acknowledge that the surrogate models used in this work are simple and limited, which we consider acceptable given that this is not the main focus of the study. We agree though that more advanced surrogate modeling approaches could be explored, and we have added a brief note mentioning a follow-up study using stochastic polynomial chaos expansion (SPCE). Neural networks were not considered due to the limited amount of available training data.

Comment:
Add sensitivity analyses on the number of turbulent seeds and the size of the training/validation sets.

Answer: While we agree that a sensitivity study would be valuable to understand the requirements for modeling the seed-to-seed uncertainty it is beyond the scope of this initial feasibility assessment. We see this as a first step for future work and have clarified this by mentioning the sensitivity study as a follow up topic in the conclusions.