

**Answers to the comments of the reviewer 2 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 2

General Comments

Comment:

In this study, the authors developed and trained two surrogate models for predicting wind turbine loads and evaluated their performance using SCADA data. The surrogate models exhibited varying levels of accuracy in load prediction. The topic of using surrogate models for turbine performance evaluation is current and relevant to ongoing research in the field. The paper is generally well-structured, with a logical and coherent flow.

Answer: We would like to thank the reviewer for this positive judgement.

Comment:

However, it does not clearly emphasize its innovations, and the interpretation of the results is not fully convincing. The methodology, including the use of PCE and GPR models and reference wind turbine model, is well-established in the literature, and thus do not present novelty to the reader.

Answer: As suggested by Reviewer 1, we have revised the introduction by

moving the discussion of generic surrogate models earlier in the text. This makes our contribution and innovation clearer in the context of established methodology. We agree, that the surrogates used are rather standard. The paper did not aim to introduce novel methods for surrogate modeling, but to investigate the use and potential of generic wind turbine models in combination with surrogates.

We have rephrased parts of the discussion and conclusions to provide a clearer interpretation of the model performance. The revised passages read as follows:

Discussion: For the blade loads, the median prediction errors were consistently within 10–15%, a level that can already be considered useful for practical applications. For the tower accelerations, the errors were notably larger, typically around 30–40%, yet still within a range that allows the models to support initial load assessments and comparative studies. At the same time, these deviations highlight that further refinement is needed before the surrogates can be used for applications requiring high quantitative accuracy.

Conclusions: For the blade loads, the surrogates achieved median relative prediction errors of around 10% for maximum loads and 10–20% for the mean loads. This accuracy suggests that the blade load models can already be applied for practical purposes such as initial load assessments. In contrast, the tower acceleration models performed notably worse, with median prediction errors typically between 30–40%, and even higher errors for maximum accelerations, limiting their current usability.

Comment:

Furthermore, the interpretation of the prediction accuracy is questionable. For example, the annual median error in tower acceleration exceeds 40%, which raises concerns not only about the accuracy of the surrogate model itself, but also about the consistency between the reference turbine used in the OpenFAST simulation and the actual on-site load measurements. This concern hinders the reproduction, generalization, and practical application of the proposed approach (surrogate + generic WT models).

Answer: Indeed, the simplifying assumptions in the simulation of the reference turbine do contribute to the deviation between the surrogate and the SCADA data. We are now reporting simulation results comparing both for selected operating conditions, quantifying roughly the order of the modeling error. The surrogate error contributes further and is now also investigated in more detail in the appendix. At the same time, even in view of the deficiencies of the surrogate and reference turbine modeling approach, errors of acceptable size for practical applications could be obtained. The approach is clearly described and reproducible, in particular with the published code at the same time we agree that the aspect of generalization requires more research with larger data sets.

Comment:

The attached comments are provided for each section of the manuscript, with the hope of improving the quality of the paper.

Answer: We are grateful for the detailed comments and improved the manuscript based on the suggestions and questions. Below, each point is addressed, including the quotation of changes done to the manuscript.

Abstract**Comment:**

The authors present a concise abstract that briefly introduces the methodologies applied in the study, and the conclusion regarding the use of surrogate models for load prediction. However, the research question is not clearly stated, leaving the reader uncertain about the specific problem the study aims to address.

Answer: This point was also made by reviewer 1. We agree with this assessment and have revised the abstract accordingly to explicitly state the problem our study addresses.

Comment:

In addition, the claim of achieving ‘reasonable accuracy’ (line 7) requires clarification, as the blade load and tower acceleration report distinguish errors. This variation should be acknowledged and discussed more explicitly.

Answer: We agree and rephrased the sentence to: **While the models approximated the real-world turbine behavior with reasonable accuracy for the blade loads, the prediction quality for tower accelerations was notably lower.**

Introduction**Comment:**

In this section, the authors identify two key challenges in aerodynamic simulations: high computational cost and limited access to detailed turbine models. Surrogate models are then proposed as a means to reduce computational effort, while the adoption of a reference wind turbine model is suggested as a potential solution to the issue of limited public model. However, the motivation or assumptions the selection of PCE and GPR models require further clarification. In addition, the literature review does not identify existing gaps in the application of surrogate models, particularly PCE or GPR, for wind turbine load prediction. As a result, the novelty or contribution of the study is not convincingly demonstrated.

Answer: We agree with the point, that the novelty of the research needed highlighting. This was also mentioned by reviewer 1 and has been addressed

above. The novelty of the work is also now stated directly in the introduction: While surrogate modeling is therefore an established approach, the novelty of this study lies in assessing the feasibility of combining it with generic reference turbine models to predict real turbine loads on a large scale, thereby reducing reliance on commercial models that are typically inaccessible. To this end, we use a database of five years of SCADA data to evaluate the accuracy of the combined surrogate-RWT approach.

Moreover, we aimed to clarify the choice for PCE and GPR in the introduction, by rephrasing the relevant passage to: Given that the turbulence-induced variance sets a natural limit on predictive accuracy, our surrogate models - based on PCE and GPR - will estimate both the mean and the standard deviation of the loads, as these approaches have been proven effective in similar applications in the literature.

Comment:

Although the authors aim to address data limitations by using a reference turbine model, the lack of either public available real turbine models or SCADA datasets makes it difficult to assess the broader significance of this study or novelty of the approach.

Answer: We agree that the lack of publicly available turbine models and SCADA datasets is a challenge for the community. The absence of real turbine models was the main motivation for our study. Furthermore, the SCADA dataset used here has been published, so our case study can be reproduced and built upon. It is correct, however, that extending the approach to other turbine types would require the publication of additional datasets, which could be facilitated by future studies.

Comment:

1) The authors state that the development of surrogate models remains cost-intensive due to the randomness associated with wind turbulence. However, the discussion around the number of seeds used in simulations is unclear. While the cited literature uses varying seed numbers (e.g., 6, 100, 8), the paper does not provide a clear conclusion on what constitutes a sufficient number of seeds for accurate load prediction. It would be valuable to clarify how the number of seeds affects the reliability or prediction accuracy of the surrogate model, and what the potential implications are of using an insufficient number, such as the recommended value of 6 by IEC standard.

Answer: In this study, we aimed to simulate as many seeds as possible, but budget and time constraints limited us to 30 realizations. Since a systematic convergence analysis on the required number of seeds requires additional data, we now state this as a point for additional research in a follow up study. Still, the effect of the limited sample size is roughly visible in Figure 4, where standard deviations due to the seed-to-seed variability are drawn; indicating a manageable effect for the blade-related quantities and a large relevance for the

tower-related outputs.

Comment:

While the PCE model used in Murica et al. (2018) required 100 realizations to predict the mean and standard deviation of the load, the authors of this study describe that approach as having a "high computational cost" (line 40). However, this study also adopts the PCE model and claims it "can overcome the computational burden" (line 40), without providing a clear explanation of the number of seeds used or the rationale for selecting the PCE model over other alternatives such as ANN, RSM, or GPR, which are mentioned briefly in lines 25–26. A more detailed justification for the model choice and a discussion of its computational trade-offs would strengthen the study, and reduce potential confusion for readers.

Answer: Thank you for pointing this out. The statement of *high computational cost* is indeed misleading in this context as formulated and has been removed. Ideally, one should formulate a target accuracy (e.g. for the mean response) and carry out as many seed-to-seed replications as needed. As argued above, due to budget and time constraints we were not able to evaluate the required size precisely. However, no matter how large the required sample size, the PCE approach is expected to provide a much faster access to this mean value over the entire parameter space, compared to purely evaluating the simulation model and this is what is meant by *can overcome the computational burden*. We didn't pursue NNs because of the small sample size, which is now mentioned in the text and chose GPR and PCE because of their known ability to work in the small-to-moderate data regime, with a moderate number of inputs. We added *"Moreover, PCE and GPR were chosen as relatively simple methods that perform well even with limited training data, making them well-suited for this initial feasibility study."* to the introduction.

Comment:

In line 44, the GPR model, described as not relying on replication, is also selected in this study, which helps to overcome the high computational burden associated with turbulence-induced randomness. However, the differences between the PCE and GPR models are not clearly presented, making the rationale behind the selection and comparison of these two models unclear to the reader. Especially, the GPR model requires careful Kernel selection and are sensitive to noise data. A clearer explanation of the distinctions between the models and the criteria for their selection would improve the clarity and justification of the methodology.

Answer: Here the GPR model was applied in the same way, as a pure regressor to approximate mean and standard deviation separately. Hence, we didn't study heteroscedastic GPs that could indeed work without replications. Hence, there is not much difference in the way the methods are applied and we decided therefore, to keep these parts short. Indeed, choosing a kernel is relevant and difficult but the chosen one is widely used and appeared to be a reasonable

choice for these first attempts.

Comment:

Line 52-53 outlines the use of RWT and surrogate models to address two key challenges. However, previous studies, for example [1] have already demonstrated that the PCE model can deliver accurate performance in turbine fatigue load prediction, with differences of only around 5% compared to high-fidelity simulations for various site-specific conditions. Furthermore, in this study, 30 seeds are used in simulation, but it remains unclear whether this setup overcomes the computational burden. Since one of the main motivations for surrogate modeling is computational efficiency, this point should be more convincingly justified. [1] Dimitrov, Nikolay, et al. "From wind to loads: wind turbine site-specific load estimation with surrogate models trained on high-fidelity load databases." Wind Energy Science 3.2 (2018): 767-790.

Answer: The main computational savings of our method, come from the fact that once the surrogate models are trained, they can be used to evaluate loads across the full SCADA dataset without the need for costly aeroelastic simulations. In this sense, the setup overcomes the computational burden mentioned. We have also added this explanation into our introduction to improve clarity. As mentioned above, the seed size should not be chosen ad-hoc but determined based on an estimated of the associated sampling error. This was unfortunately not feasible in our study due to the aforementioned constraints.

Section 3: Case Study

Comment:

Please clarify 'parameter B away 1' (Line 229).

Answer: Thank you for pointing this out. We changed this to: **However, due to the absence of mast wind measurements, there is no empirical basis to calibrate the parameter B . Therefore, it is kept at a value of 1 in this study.**

Comment:

Line 240 mentions that simulations were performed to verify the absence of critical resonances under operational scenarios. While this test helps confirm the basic functionality of the turbine model, it does not substitute for validation against real-world turbine model, given that minimal adjustments were made to the RWT model (as stated in lines 238-239). Therefore, a comparison between simulation results and on-site load measurements under similar environmental conditions is recommended. This comparison would help to demonstrate that the turbine model in simulation can predict comparable load as actual turbine.

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

model results, with the aim of estimating the deviation between the adapted reference turbine model and our on site measurements.

Comment:

Meanwhile, this study applies SCADA data to evaluate model accuracy. However, two sources of uncertainty remain: the prediction errors may originate from either the surrogate model or the RWT model used for simulation. As a result, the interpretation of prediction accuracy is unclear.

Answer: It is correct that we are dealing with different sources for uncertainty, which are contributing to the observed prediction errors: the surrogate models themselves and the underlying generic turbine model used in the simulations. Given that the applied surrogates showed good fits for the mean responses (except for the towers maximum acceleration) - while larger discrepancies only appeared in the case study - it is likely that the dominant contribution to the overall error stems from the generic turbine representation and the limited set of input parameters considered. However, while a quantitative separation of these contributions is difficult and remains an open task for future work, we integrated an overview of uncertainties in the modeling workflow to increase transparency.

Comment:

In Section 3.4, Line 254 states that 30 seeds were used in the simulation, line 255 explains it as a compromise between small and large sample sizes, but no further justification or explanation is provided for this choice. As discussed in the introduction, the literature review in the manuscript does not establish a clear standard or conclusion for the appropriate number of seeds in such simulations. Given that one of the study's primary motivations is to address computational burden, it is recommended to perform a sensitivity analysis on the number of seeds. This would help assess the impact of wind modeling variability and demonstrate whether using 30 seeds is sufficient to capture the inherent stochasticity of wind conditions, thereby supporting the reliability of the results.

Answer: As pointed out above, the number of seeds does not aim to reduce the computational burden of our proposed method, since the main savings are obtained in the application of the surrogates. We agree however, that a sensitivity study on the number of seeds, should be conducted in a follow up project. To clarify, we revised the sentence concerning the choice for the number of seeds: *This seed number aligns with TurbSim's documentation [1] as a compromise between the small and large sample sizes found in the literature. While a larger number of seeds would be desirable, with the required number depending on the desired statistical accuracy, the choice of 30 represents a practical limit for this first feasibility study.*

Comment:

4) Line 262 mentions the duration of simulation, but the reason for this specific 630-s is not explained. It is unclear whether 30 s are sufficient to mitigate the initial transient effect, considering the wide speed range. This is important for dynamic responses and load predictions. Further justification, such as signal stabilization plots or references, are suggested to demonstrate.

Answer: This was also pointed out by the reviewer 1 and references are now provided.

Section 4: Results

Comment:

Table 3 compares the prediction errors between the PCE and GPR surrogate models. As the mean MAE and RMSE for mean blade moment and std of tower acceleration are within 10%, but their std are much enlarged. Line 300 states that the response surfaces of the std are less smooth, resulting in higher errors for these values. However, the reasons for the coarser surfaces are not discussed. More clarifications are suggested.

Answer: We agree with this comment and have expanded the explanation in Section 4.1.1 to clarify the potential reasons for the less smooth response surfaces. The revised text reads:

This lack of smoothness may arise because the estimates of the standard deviations have not yet converged with only 30 turbulent realizations, or because the underlying surfaces are inherently less smooth. Using a larger dataset with more realizations could help clarify this and potentially reduce the observed errors.

Comment:

Also, the influence of model parameters, such as the choice of polynomial order in PCE or kernel type and hyperparameters in GPR, on the accuracy should be addressed. A sensitivity study would help clarify the source of the observed discrepancies. Given the GPR model, which is free from seed number effect, provides similar or slightly lower std values for blade moment. Simply suggesting that a larger seed number would potentially improve smoothness (line 300), is not a sufficient justification without supporting evidence.

Answer: We would like to clarify that both the PCE and GPR models rely on the 30 turbulent realization in order to estimate the mean and standard deviation of the target responses. The impact of polynomial order and regressor is addressed in the presentation of the training results. However, we did not conduct a sensitivity study on the kernel choice for GPR. Fitting a Matern kernel would be a logical next step, including also sample paths with reduced regularity. At the same time, we expect that improving the model itself promises more gains in accuracy towards the recorded data. An analysis of the convergence

behavior of our models is now integrated in the appendix.

Comment:

This concern also applies to the statement regarding the maximum values due to small parameter variations, where increased seed number may not improve the prediction accuracy. Such claims should be supported by quantitative analysis or validation to avoid speculative conclusions.

Answer: We agree and have now removed this speculative statement.

Comment:

Figure 3 compares measured data, openfast simulation results, and predictions by surrogate models. However, the markers appear densely clustered in each subplot, making it a bit difficult to clearly distinguish among different categories.

Answer: The figure has been reworked by splitting it into two subfigures and reducing the number of plotted points.

Comment:

Moreover, the importance of seed uncertainty mentioned in line 332 is not fully reflected or supported by the figure, as the simulation model differs from the real turbine used in measurements, and this impact of model uncertainty is not clearly quantified. This uncertainty further limits the evaluation of seed effect.

Answer: The importance of seed-to-seed uncertainty is derived directly from the 30 replicated simulations, as reflected by the error bars in the figure. While the absolute magnitude may differ for the real turbine, we expect the present and qualitative relevance of this uncertainty to transfer to the real turbine.

Comment:

To improve clarity, it is recommended to separate the comparison into two figures: one comparing measured data with surrogate model predictions, and another comparing measured data with OpenFAST simulation results. This separation would provide a clearer overview of the accuracy, and reliability of both the simulation and surrogate models.

Answer: To improve clarity, we have separated the results into two figure and reduced the amount of sampled data so the plots are less crowded.

Comment:

Line 355 states that the annual prediction errors could potentially reflect aging trends of the turbine. However, this statement is potentially misleading, as there is no justification provided to confirm that environmental conditions, such as wind speed and turbulence intensity, were comparable between 2017 and 2022. Without such validation, it is difficult to attribute changes in prediction error solely to turbine aging.

Answer: As recommended, we added two figures illustrating the annual distributions of wind speed and turbulence intensity and have refined the accompanying analysis.

Comment:

A similar concern applies to the statement in Line 360 regarding the role of the pitch controller in extending turbine lifetime. To strengthen these claims, it is recommended to include a comparison of key environmental parameters (e.g., turbulence intensity, wind speed distribution) across the five-year period. In addition, prediction errors should exclude the model uncertainty, as discussed. This would help isolate the influence of external factors, and focus on the ageing impact.

Answer: We agree that a separation of model uncertainty and external influences would be useful. However, given the scope of this feasibility study, it was not possible to isolate model uncertainty quantitatively. The concern about a comparison of the mentioned parameter has been addressed above.

Section 5: Discussion

Comment:

The statement regarding the effectiveness and efficiency of the surrogate models in Line 375 is not fully supported by the results presented. While Table 3 shows that the trained PCE and GPR models achieve prediction errors below 10% for the mean blade root moment and the standard deviation of tower acceleration, the errors exceed 10% and even 20%, for the standard deviation of blade root moment and mean tower acceleration, respectively. These discrepancies suggest that the surrogate models seem less reliable for capturing variability than central tendencies.

Answer: We now highlight this observation more directly in our manuscript: **Concerning the developed surrogate models, it can be noted that both the PCE and GPR were generally effective and efficient at capturing the mean responses of the training data, but were less successful for the standard deviations, and no satisfactory fits could be obtained for the maximum tower accelerations.**

Comment:

Furthermore, the claim of efficiency, based on the use of 30 seeds, is not explicitly validated. No sensitivity analysis is provided to assess how the number of seeds influences prediction accuracy, making it difficult to conclude that the chosen setup is computationally efficient or statistically sufficient.

Answer: As discussed above the choice of 30 turbulent seeds does not aim to reduce computational burden. We agree that it represent a limiting factor in how accurately we can capture the seed to seed uncertainty. Concerning the computational efficiency, we consider explicit validation unnecessary, since applying surrogates to the full SCADA dataset is by design far more efficient than simulating every operating point with the aeroelastic code.

Comment:

The uncertainties arising from both the wind turbine model used in OpenFAST and the surrogate models are not independently identified or quantified. Without a clear separation of these sources, it is difficult to determine the contribution of each to the overall prediction error. Therefore, the claim in Line 380 lacks sufficient support. A more rigorous uncertainty analysis, distinguishing between model structural errors and surrogate approximation errors, would be necessary to justify it.

Answer: To clarify, at this point in the text, our statement does not refer to predictive accuracy but to the practical feasibility of developing surrogate models in an industrial context. Later on in the discussion we talk about how usable the approach seems, with the observed prediction errors in mind. Concerning the separation of uncertainties, we agree that it would be a valuable next step. However, we are of the opinion that, when comparing the relatively low surrogate validation errors (of the mean responses) with the larger prediction errors observed in the case study, most of the discrepancy is likely attributable to the turbine model and its simulation rather than the surrogates themselves.

Comment:

The limitation mentioned in Line 389-390, regarding model controller actions or system constraints, is not clearly demonstrated or reflected in the presented results. This claim appears disconnected from the analysis or result, and would benefit from further clarification.

Answer: We agree with this comment and have revised the sentence: **This indicates that the approach cannot fully account for changing operating conditions, whether driven by environmental factors or technical modifications, due to modeling constraints and the limited availability of detailed information.**

Comment:

As state in lines 394–399, the simplified process limits the generalizability of this study. However, given the prediction errors observed between the surrogate models and the simulation model, it is unconvincing to conclude that simplification alone has the greatest impact on prediction accuracy compared to measurement data. Other factors, such as model assumptions, data quality, and inherent uncertainties, should also be considered and discussed to provide a more comprehensive understanding of the sources of error.

Answer: We agree with this assessment and have adjusted the manuscript to be more balanced: This simplification is expected to have had a major impact on the prediction accuracy achieved in the presented case study. However, other factors, such as the limitation to four input features, their calculation (e.g. in the case of the air density), the relatively coarse grid for the simulation of the wind field or data quality issues in the SCADA dataset, also need to be considered as contributing sources. A detailed overview of expected uncertainties in our modeling workflow is presented in the Appendix.