# General comments:

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. 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.

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).

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

## Abstract:

1) 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. 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.

## Introduction:

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.

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.

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.

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.

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.

2) 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.

## 3-case study:

1) Please clarify ‘parameter B away 1’ (Line 229).
2) 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.

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.

3) 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.

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.

## 4-Results

1) 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.

   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.

   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.

2) 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.
   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.

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.

3) 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.

   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.

## 5-Discussion

1) 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.
   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.

2) 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.

3) 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.

4) 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.