General Comments

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. 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.

General Writing

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

Abstract

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.

Introduction

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

Section 2: Surrogate Modelling

- Line 100: Please elaborate on the methodological extension related to the pitch angle.
- Line 121: The reference to Dimitrov et al. (2018) is abrupt. Adding more context would make it smoother.
- Line 124: the reference to Mara and Becker should also be introduced with more context.
- 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.

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

Section 3: Case Study

- Line 205: Please clarify the limits under which generic turbine models can be considered applicable to real-world scenarios.
- Line 224: citations to be improved.
- 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.
- 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.
- Line 258: Indicate the OpenFAST version used.
- Line 264: Justify the use of 30 seconds as transient time with relevant references.
- 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.

Section 4: Results

- 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.
- 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.
- Line 304: Clarify what the “respective other model” refers to.
- Line 307: Specify which convergence analysis is being referenced.
- Line 310: Provide the mentioned error curves in the manuscript.
- 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. 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.

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:

- Quantify the error between generic model outputs under measured environmental conditions and actual measurements.
- 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.
- Perform a convergence study on the number of seeds to take into account to capture the seed-to-seed uncertainty.
- Discuss other surrogate models, such as neural networks, which have shown promising results in similar contexts.
- Address the poor performance at low wind speeds, and elaborate on way to address it, expanding the dataset could be a way.
- Tune the hyperparameters of the models using grid search or random search for instance.

Section 5: Discussion

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

Summary and Recommendations

The scope of work is highly relevant, and the preliminary findings on turbine aging and inter-turbine variability are promising. However, substantial improvements are required before publication. Key recommendations are:

1. Provide an error estimation from discrepancies between the generic OpenFAST model and measurements.
2. See the impact of increasing the size of the training dataset on the model performances.
3. Elaborate on surrogate model hyperparameter tuning, considering grid search or random search.
4. Expand the discussion on alternative models tested, consider implementing neural networks with larger datasets.
5. Add sensitivity analyses on the number of turbulent seeds and the size of the training/validation sets.