May 31, 2024

Review of the article:
Probabilistic surrogate modeling of damage equivalent loads on onshore and offshore wind turbines using mixture density networks

General comments:

The article addresses the problem of probabilistic surrogates for aeroelastic wind turbine models, a relevant scientific question within the scope of WES and with significant impact on the wind energy industry. The article includes interesting new ideas and techniques in the use of mixture density networks to predict the local distribution of the response of fatigue loads given a set of turbulent wind inflow and sea state parameters.

The article is well structured and written concisely. The scientific methods, analyses and assumptions are valid and clearly outlined and reproducible. The presented results (convergence study with number of aeroelastic simulations, with two out-of-sample test datasets) are sufficient to support the interpretation, discussion and conclusions.

The authors cite sufficient and relevant work in their literature synthesis.

Some points to improve:

- The objectives of the study or hypotheses should be stated in the introduction.
- The original contribution of the paper to the field should be clearly stated in the introduction.

Specific comments:

- **L20**: You are missing a general description and references to the most used aeroelastic models: OpenFAST, HAWC2, Bladed, Flex, etc.
- **L32**: The reference to "fast frequency-domain reduced-order models" seems disconnected from the article. It should be either removed, or explained why these techniques are interesting. If these techniques are interesting, why not testing them in the article? Are these ROMs a way to replace the aeroelastic model (OpenFast)? Are the probabilistic surrogates presented in the article relevant for ROMs. I do not necessarily agree to describe ROMs as physics-based models, as they can be trained on simulations just as your load surrogates. Are aeroelastic models not physics-based?
- **L42**: The uncertainty due to the presence of unknown or inexpressible features is an epistemic uncertainty. Epistemic uncertainties represent lack of knowledge and therefore they could be reduced by for example: improving the accuracy of a measurement, or increasing the number of degrees of freedom (DOF) used to characterize a phenomena. If you represented the uncertainty in the turbulent-wind and sea-state field with more degrees of freedom (let's say a grid with three wind/wave components) you would not have this uncertainty. Since you use few integrated quantities to represent the inflow you have the epistemic uncertainty of the flow realization.
- **L57**: TurbSim is not a turbulence model, but a software. You should refer to the turbulence model used, i.e. Mann? KAimal spectra? etc. You should also report the parameters used on the spectra.
- **L68**: Deterministic models can be used to build probabilistic surrogates similar to the ones you propose: Quantile regression proposes the use of multiple deterministic surrogates to

predict the local distribution (Koenker, Roger, and Kevin F. Hallock. 2001. "Quantile Regression." Journal of Economic Perspectives, 15 (4): 143-156. DOI: 10.1257/jep.15.4.143) (Meinshausen, N. and Ridgeway, G., 2006. Quantile regression forests. Journal of machine learning research, 7(6).). You should add a sentence here on this. Note that these types of probabilistic surrogates need to have multiple seeds in the training in order to compute the local quantiles.

- **Figure 1**: The deterministic model should also have the explicit dependency on the inputs $\hat{y}(\mathbf{x})$.
- **L96**: This sentence needs to be re-written for more clarity. These probabilistic models can be train without having to simulate multiple inflows for each input vector, because the methods can infer the local distribution based on the samples available in the neighborhood. This point is later clarified in L240-L242. But is should be clear from the beginning.
- **L265**: Missing citation to Sobol.
- **L270**: The distribution of wave state parameters is fitted using KDE based on what? The following text and Table 2 are confusing because you focus on the lower and upper bounds, but this is not relevant for the KDE.
- **Table 2**: The distribution of the turbulent wind field parameters is reported to be uniform in table 2, but seems not to be the case on the marginal distributions on Figure 3. This table should be accurate description of the simulations.
- **Table 2**: Why use a random uniform distribution of seeds and not just an increasing seed number? Consecutive seed numbers will provide a similarly independent field as "random" seeds.
- **Table 3**: The symbol Δ needs to be defined as the discretization step.
- **Figure 3**: The marginal distributions should be presented as histograms and not as KDEs as it gives a wrong sense of the actual distribution.
- **Figure 3**: The marginal for the shear exponent should be re-scaled to show the distribution of the train and test 1, instead of focusing on test 2.
- **Figure 4**: The outliers at high wind speed that give low moments (and that cause the bi-modal local distributions) should be highlighted and explained. Why do they occur?
- **L336**: It is not clear why you are using the standard deviation instead of the damage equivalent load for the convergence study. I expect the DEL to be harder because of the power exponent. This should be corrected.
- **Table 4**: The label should explicitly state that these are R^2 results.
- **L407**: The description of the reference local distribution should be given in the methodology, i.e. table 3. and not here.
- **L418**: The conclusion of the causes of the low performance at low TI should be illustrated by showing the local distributions, as you did in figures 11 or 12.