

## **Manuscript: The Influence of Flow-Field Representation on Electrical Power and Fatigue-Load Surrogate Modelling in Offshore Wind Farms**

This manuscript is a revision of the paper “On the Sufficiency of Low-Order Spatial Descriptors for Fatigue Load and Power Prediction from Aggregated Rotor-Plane Flow Fields in Offshore Wind Farms”. However, the paper was significantly restructured, the applied methods were changed, and the focus was shifted. Hence, it is more or less a new paper, which is the reason why my review does not address the previous points but raises new ones.

The revision now investigates how different surrogate models based on different (spatial) information on wind speeds and turbulences across the rotor disk perform in calculating fatigue loads and electrical power of offshore wind turbines in a wind farm. The previous version of the paper proposed to use spatially aggregated information, e.g., mean values, shear gradients, wake center, etc., to train surrogate models. This approach is no longer proposed. Instead, three other approaches are presented. 1) The full spatial data is directly used in an CNN. 2) An MLP is trained using full data, but the spatial information is neglected. 3) First, a PCA is applied to the data to reduce the number of inputs. Second, the outputs of the PCA are used to train an MLP.

The relevance and the innovation of the paper is low. There are already several surrogate modelling approaches for wind turbine fatigue loads or electrical power. Further development is not clearly shown. The mentioned research gap (l. 45) of analyzing both (loads and power) in a single paper exists. However, for me, it is not clear why this should be advantageous. It is not surprising that different surrogate models are most suitable for different outputs (here: fatigue load and electrical power). A second very problematic point is that it is not clear what a potential application could be. For a classical design application according to the standards, spatial information is not available. Hence, these surrogates cannot be applied. Spatial information is only provided by aero-elastic simulations (or measurements). And if you have to already run thousands of aero-elastic simulations to train the surrogate for what purpose do you use the surrogate?

Hence, I cannot recommend this manuscript for publication in WES.

Some additional relevant shortcomings are listed below:

- 1) Since different surrogate models are used (CNN and MLP), it is not clear whether the spatial representation is the actual reason for the differences. In my opinion, you mainly compare different surrogates and not different spatial representations.
- 2) Why is it important to analyze surrogates for fatigue loads and electrical power jointly?
- 3) Considered loads are not fully clear, e.g., the coordinate systems are not defined, so that bending loads for the main bearing are not defined, and tower top and tower bottom seem to have different coordinate systems.
- 4) If I am not mistaken, you train your surrogates in each fold with  $800 \times 4/5 \times 6 \times 48 = 184.320$  samples. This is a lot of data. As each sample is based on an aero-elastic simulation, such a high amount of training data is not practical. For standard surrogates (without spatial information), normally only a few hundred to a few thousand samples are used.  
Furthermore, it is not investigated how much training data is required for each surrogate. This convergence behavior is more important than the final accuracy for a high number of training samples. Final accuracy is normally quite high for all types of (advanced) surrogates. However, reducing the number of training samples required is essential for real applications.
- 5) Section 5 requires more explanations and reasoning. For most of the section, it is not clear why you have chosen which surrogate/hyperparameter/architecture/etc. Some choices seem arbitrary.

- 6) If I see it correctly, the errors in power production are really low (NRMSE of 1% or less). Hence, arguing that one surrogate is better than another is not expedient. Here, more relevant is the amount of required training data (see remark 4).
- 7) A comparison with results of a “simple” surrogate using only fully aggregated spatial information, e.g., only mean wind speed and variance (both spatially averaged) or only the wind speed and variance at hub height, is missing and would be very beneficial to demonstrate the relevance of spatial information.
- 8) The results of the parameter study of Section 8 are used in Section 6 and 7. Hence, the order of the sections is not logical.
- 9) Limitations of this work and a critical discussion regarding the relevance are missing.