The paper titled "*System identification of offshore wind turbines for model updating and validation using field measurements*" is well-performed multidisciplinary research focusing on two main concepts:

1) Tunning and validating generic offshore turbine model using condition monitoring system and SCADA

2) Estimating rolling contact fatigue of pitch bearings

Both above concepts are relevant to the current developments in wind industry and add high value by answering important questions specifically relevant for offshore lifetime extension where model accuracy becomes more important. The topic is becoming more crucial with first offshore wind farms reaching end of design life.

The work also includes several novel aspects including:

- Combine OMA with physics-based system knowledge to improve accuracy, including distinguishing closely spaced modes and filtering deterministic load and control effects when using this technique.

- Bridging between the measurements and analytical models for estimating lifetime of the bearing and study of the effect of environmental parameters on the damage consumption.

- Adding degrees of freedom in modelling the shaft to reduce the errors in estimating the modal parameters

The work is worth publishing in the Journal of Wind Energy Science after few modifications listed below.

Major comments:

1) The manuscript would benefit from clearer boundaries between the two main topics (model development and bearing lifetime assessment), each of which could stand alone as independent research. Please make this distinction explicit—particularly in the abstract and introduction.

2) In the second part of the research, the wind speed distribution of the site is used while keeping TI of the class for defining the Rolling contact fatigue (RCF) lifetime of the pitch bearing. The rationale for this choice is unclear. If the aim is a site-specific assessment, why not use the site's measured turbulence? If the aim is solely seeing the effect of inputs in each class, why not fully commit to class input levels? Please make the justification explicit in the manuscript.

3) The study of the effects of environmental parameters on the fatigue life of the bearing needs stronger base and elaborations for the reader. While the results are valuable, they do not seem to be continuation of the flow of the study (site-specific versus design for the case study turbine using the tunned and validated model would have been more connected). Please consider making a flow through explanations in the text or justifications based on the response to comment #2 above.

4) Suggestion: As the full measurement of environmental parameters are in hand, for the OMA purpose, the responses in standstill state can be filtered to include mostly durations with low turbulence. This way the condition could have been closer to the white noise assumption in OMA. Please either apply or comment.

5) The validations should have been based on load response and/or other model outputs instead of just power curve. Although other parameters (pitch and rpm) are mentioned as considered, only presenting the power curve does not add high value. Please consider validating against a parameter which is not obtained already in the scaling step. If the concern is maintaining confidentiality, normalizing and/or tabular presentation of percentage differences can be helpful. Please also consider that investigating the scatter of output (variance) with scatter of input (e.g. high and low tail of turbulence in each MWS) is important (not only looking at the mean output and mean TI as stated in the paper).

   Suggestion: It is understood that the strain gauges are not calibrated but a track of fatigue load responses via measurements under two specific operational conditions (in which the controller is active) followed by doing the same using the aeroelastic tunned model can show whether the model shows a similar ratio/regression in the response. This shows the validity of the model to be used for fatigue load comparisons under different environmental inputs (which is common is assessing lifetime extension).

   If you skip doing the suggestion, you should mention the load validation as a missing point and suggestion for future work.

6) This type of work (especially tuning of the generic models) includes a lot of uncertainty based on different assumptions and the missing information. A discussion section maneuvering on different aspects (e.g. scaling method, assumptions, methods used and measurements) are their possible effects that are needed and will add value. This is especially important if presenting other variables for validation (see major comment #5) remain missing.

   Suggestion: The study '<u>Robertson, Amy N., et al. '*Sensitivity analysis of the effect of wind characteristics and turbine properties on wind turbine loads.*', Wind Energy Science 4.3 (2019): 479-513</u>.' can be a good base for finding areas to focus on in the discussion regarding model bias or uncertainties and their effect in the current work. The applicability of the model for site-specific fatigue assessment is a better match for consideration in the current paper (due to lifetime extension and fatigue assessment of the pitch bearing).

7) The effects of turbulence can be considered predictive as more turbulence often brings more control (pitch) activities and thus more contact rolling fatigue, no? Please consider the discussions.

Minor comments:

1) Overall, the paper is wordy, particularly in the introduction, which has room for more briefness. Throughout the manuscript, it is difficult to keep track of both the big picture and the detailed points at the same time. To improve readability and flow, I suggest restructuring: consider

placing all high-level descriptions in one section (e.g., Methodology) and then presenting the detailed assessments in the subsequent Results section. This may also address the concern raised in Major Comment #1 as it can provide more focus in introduction on the flow of the work.

2) The reason for the 1st in-plane blade mode not being detected via high frequency measurements while out-of-plane is detected is worth discussing in the paper. Is it the orientation of the sensors? Is there any correlation between having mostly turbulence related excitation in T1 turbine location during stand-still? Or any other reasons? Please consider explaining in the text.

3) Please list all the design parameters available at the beginning of section 2. This information is highly important while floating in different places in the text. Consider listing or even making a table introducing them.

4) Section 2.1.3: Is the actual profile of the turbine in hand? This needs a bit more clarity in wording to understand

5) Omitting the word edgewise and flapwise and sticking to in- and out- of plane will improve the readability and clarity. This can be clarified once for the whole paper (when describing the 90 degrees pitch).

6) Please elaborate on the measurement tools used for environmental data collection as well as their placement (in relation to each of the four turbines).

7) Line 439: the effect of mean wind speed is worth mentioning.

Technical corrections:

1) Please correct the typos below:

   Line 92: *"eingenfrequencies"* → "eigenfrequencies"
   Line 221: *"Similary"* → "Similarly"
   Line 278: *"baldes"* → "blades"
   In few places in the document (please search and correct all): *"asymetric"* → "asymmetric"
   Line 288: *"suDyn"* → "SubDyn"
   Line 191: *"Utlising"* → "Utilising"
   Line 384: *"charachtersitics"* → "characteristics"

2) Please delete the extra **'the'** in line 292

3) Line 231: please reword below sentence for more clarity and briefness:

‘*Showing mainly in radial sensors, the mode at 5a is identified as the 2nd symmetric in-plane rotor mode while the modes at 5b and 5c are identified as the 2nd asymmetric in-plane rotor mode modes 1 and 2 respectively*.’

4) Line 213: The phrase ‘*at least three poles are identified as stable in consecutive model orders*’ is not clear.