The following shows the comments of the reviewer. The author's response to each comment is in blue. Changes that are also made in the paper are highlighted in bold letters.

General comments

The paper presents a methodology to implement raceway damage in a finite-element model of a double-row four-point contact ball bearing. The approach uses nonlinear spring elements together with pretension elements to represent the gaps caused by raceway damage. The model is compared with experimental strain measurements from two downscaled blade bearings. The subject is relevant to wind turbine blade-bearing condition assessment, and using experimentally measured damage states for the FE model is beneficial.

There are some general notes as follows:

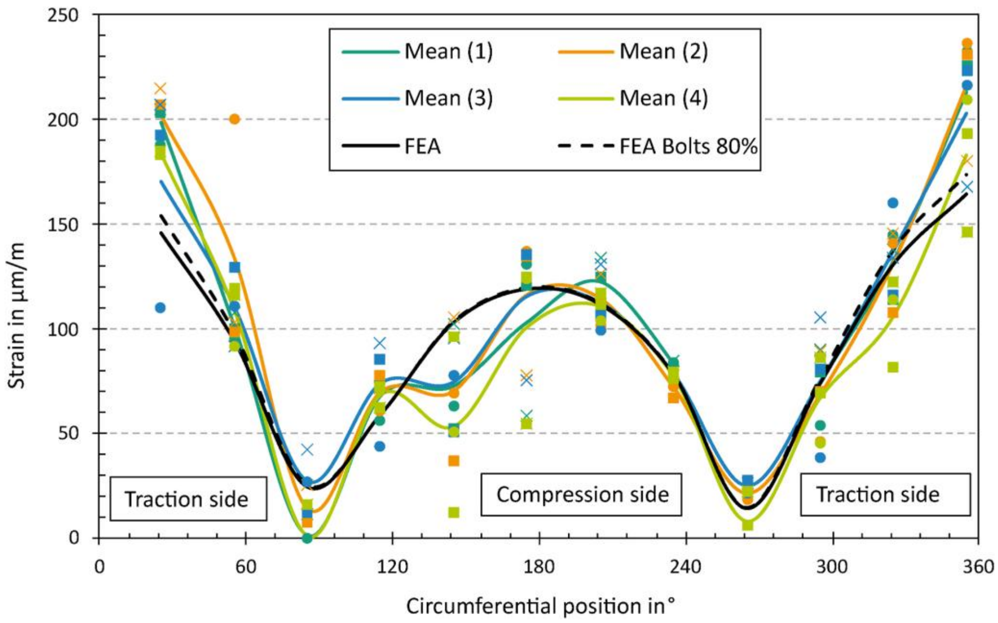
- The damage implementation is based on some simplifying assumptions. A segment is considered damaged when at least 50% of the corresponding raceway area contains material outbreaks, and a representative damage depth is used to define the gap in the FE model. Please explain the basis for selecting the 50% criterion and how representative the selected damage depths are of the measured damage topology. It would also be useful to discuss the sensitivity of the calculated load distribution to these assumptions.

  This is very good point. Internal investigations embedded in a master thesis investigated the influence of different criteria. The first approach in the thesis was a more aggressive one where no matter how small, damage was considered in the FE model. The resulting strain gauge values showed too drastic changes in the deformation characteristics. Using the 50% approach resulted in significantly better results compared to experimental data. In addition, in the tests the balls roll over the raceway and depending on the damaged size there will always be a discrete position where the ball is not aligned with the damaged part of the raceway. In the presented approach that cannot be implemented in the FE model yet. In the future we are thinking about an approach that somehow measures the damaged share of each segment to have a more detailed damage implementation in the FE model, maybe even considering different pitch angles. The published approach is the first step, and its aim is more of a general approach that is not tailored to a specific bearing, that is why we went with 50% as a compromise with an overall good fit of resulting strain gauge values.

  In total over 10 detailed measurements of damage profiles of different locations and at different stages during the fatigue tests were conducted. With an accuracy of +/-10% all results show a first plateau or multiple plateaus where arguably the ball would be in contact with the raceway again around 200 µm for bearing 0411 and 150 µm for bearing 0498. Again, it should be mentioned that this is a simplification and future FE models could represent damage in more detail.

- The validation approach is not clear to me. The FE result is considered to have zero deviation when it falls within the minimum and maximum experimental values. Therefore, a deviation of 0% does not necessarily indicate a close agreement between the FE and experimental results. Moreover, according to the second validation criterion defined in the Introduction, the characteristic course of the FE results should match the experimental mean values. Please explain how the validation approach used in Tables 2 and 3 is consistent with this criterion.

Here I would like to refer to an earlier publication ("Validation of a finite-element model of a wind turbine blade bearing", DOI: https://doi.org/10.1016/j.finel.2023.103957) where the same bearing dimensions were tested on the same test rig with focus on model validation using strain gauges. The following figure is from that publication and shows results from 4 different tests.



The points, square and crosses show the strain during the experiment for different pitch angles and the lines show the mean values over all pitch movements. It shows that the experimental results can scatter a lot while the mean values remain comparable. This highlights that every model validation compares one specific solution of an ideal FE model with a range of imperfect experimental results. The authors argue that the validation of a FE model is successful when the results fit within the range of experimental data.

Because of that, in this paper the authors always give the minimum, maximum and mean values of the test to use that range of experimental data for the validation. This range includes all imperfections. When the results of the ideal FE model now fit into this range the deviation is considered as 0. When the reviewer has ideas for different criteria for model validation, please get in contact with the corresponding author. We are happy to dive into a more detailed discussion on that even beyond this review.

The second criteria cannot be evaluated using the numbers in Table 2 and 3. The second criteria is more visual validation where the diagrams need to be checked whether the maximum and minimum simulative and experimental strain occur at the same circumferential position.

- For bearing 0498, an optimum system preload of 60 µm is determined through an iterative process, and the resulting FE model is then validated against the experimental strain data. It is not clear how the optimum preload was determined and which experimental data were used in this process. If the strain measurements used for validation were also used to determine the preload, the comparison cannot be considered fully independent validation. Please clarify the procedure used to determine the 60 µm preload and its relation to the validation data.

  The reviewer is right; this is not a fully independent validation. As shown in Figure 13, the strain gauge data is used to compare the resulting strain with and without preload. The preload due to the fit of tolerances of the raceways curvatures and ball diameters cannot be measured directly and therefore the strain gauges are used. The authors decided to implement this bearing in the paper anyway to also show the possible influence of higher preloaded bearings and the influence of raceway damage with a bending moment. As bearing 0411 is loaded with an axial force and only bearing 0498 is loaded with a bending moment which is more relevant for a blade bearing. Further, the individual preload of the ball-raceway interaction will be different. Using just one value as in this paper, always is a simplification.

- The interpretation of the calculated L_{10} life should be consistent throughout the paper. In Section 2.5, it is clearly stated that the life calculation for an already damaged bearing represents the likelihood of new damage initiation on the still intact raceway due to the altered global load distribution. However, in Section 3.3 the calculated life is referred to as the “remaining useful life”. Since the calculation does not directly predict the remaining life of the already damaged bearing, I recommend avoiding this terminology and using the interpretation given in Section 2.5 throughout the paper.

  The comment on the remaining service life in Section 3.3 is meant to state that the calculated bearing life during the disassemblies shows an increased likelihood of new damage over time which will affect the remaining service life negatively because new damages will occur. The reviewer is right that the remaining service life is not calculated in this paper and it is important to clarify that. The statement in the paper will be changed as follows (changes in bold):

The **life** calculations **that show an increased likelihood of new damage over the disassemblies** therefore indicate that the altered global load distribution has a negative effect on the remaining useful life: As the rating life goes down significantly, new damages are much more likely to form on the raceway due to the changed global load distribution, neglecting other effects.

Specific comments

1. **Section 2.3:** The bearing 0411 was disassembled and reassembled several times to document the damage progression. The authors mention that the grease was reused to maintain comparable lubrication conditions. However, repeated disassembly and reassembly may also affect the mounting conditions and bearing preload. Please explain how comparable mounting conditions were ensured after each reassembly and whether their possible influence on the strain measurements was considered.

   That is a very good point and probably one of the largest uncertainties when testing bearings in general. To minimize the risk of changes during that procedure, we have strict protocols how to handle the bearing during the evaluation process including opening and closing as well as mounting the bearing in the test rig. The procedure of tightening the bolts (sequence and load) is prescribed and kept the same for each bearing. In addition, the same two technicians handle the whole process to minimize human influence.

   Further, the showed Figure above from a previous publication investigate the influence of different bolt preloads. It can be seen that a bolt preload of 80% has significantly less influence on the resulting strain compared to deviations within the experimental data.

2. **Section 3.3 / Figure 15:** The increase in calculated L_{10} life from disassembly 4 to 5 is explained by increased bearing deformation causing balls in damaged areas to carry load again. Since this is a rather important and counterintuitive result, it is recommended to provide quantitative evidence of the corresponding bearing deformation or gap closure to support this explanation.

   The ball forces displayed in Figure 14 show that from the 4th disassembly on springs in damaged areas start to carry load again. This is significantly more pronounced for the 5th disassembly (see yellow and purple lines around 300°). That shows that the global bearing deformation is large enough that balls get back in contact with damaged parts of the raceway. It should be mentioned that any surface changes are not represented, and hertzian contacts are assumed. It remains unclear if the same effect would occur when the correct ball raceway at damaged parts would be considered.

3. **Section 3.3:** The authors state that the local surface effects neglected in the model would have an even more detrimental effect on the bearing life. Since these effects are not investigated in the present study, please provide a reference supporting this statement or formulate it more carefully.

   The sentence will be changed in the paper to have a more careful formulation (changes in bold):

   This effect is only of a brief duration, however, as disassembly 6 obtains the lowest of all lives, despite the fact that local surface changes are neglected in this calculation, which **~~would~~ could** have an even more detrimental effect on the life of the bearing **due to possible pressure peaks**.