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.

The article discusses a finite element model of raceway spalling in wind turbine pitch bearings accounting for a gap of defined (constant) depth, which then determines the change in local ball-raceway load distribution. The changes in load distribution are then used to assess the change in rolling contact fatigue life for an equivalently loaded bearing. Model predictions are compared to accelerated load tests. I've not seen this type of modeling done before, especially when in conjunction with detailed tests at a relevant scale, so I enjoyed it. However, I have a few general and several specific comments and questions for the authors to consider. Generally speaking, I recommend the authors provide additional discussion of most of the results, as this felt relatively brief and I was left with many questions.

2.3 Raceway fatigue tests

- Lines 163 to 170: I wasn't quite clear on the sequence of the test. Is it a correct understanding that the blade was pitched in pure compression (for an unstated period of time – how long?), then disassembled, then reassembled, then reference tested with a pure moment (was it just one cycle, or also for a period of time?). This process was then repeated 5 more times. Was the amount of time between each disassembly the same? A few more details like this would be helpful.

  Yes, this understanding is correct. The rolling contact fatigue test for bearing 0411 is done under pure axial compression load. The period of time for each test between dissasemblies is different due test activities on the same test rig in other projects. However this is not an issue for tests, as the time and revolutions for each period is well documented and the tests do not attempt to represent any explicit damage but focusses in the damage progression between the dissasemblies. The bending moment after each reassemble was used for the strain gauge comparison as larger deviations are shown with bending moments compared to pure axial loads and changes in the bearing behavior is more pronounced. The pitch movement under bending moment was one pitch cycle from 0° to -60° and back. This is stated in the paper in line 169/170.

  The following statement will be added to the paper for further clarification: **The amount of time and therefore rotations of each individual test are not consistent but depends on test activities on the same test rig in other projects.**

- Figure 6 and Figure 7: At first, I got a little bit confused regarding inner and outer rings and blade side and hub side relative to both figures. For clarity, I believe it would be helpful to mark "blade side" and "hub side" on Figure 6, while on Figure

7 I believe it would be helpful to state in the caption that it is for the inner ring. That is, only the inner ring is shown in each as it was more damaged than the outer ring. At least, if that is the correct understanding. Similar for Figures 8 and 9.

The blade side in Figures 6 and 9 will be labeled as BS and the hub side as HS. It will also be mentioned in the text. Figure 7 actually shows the damage for the diagonals 1 and 3 (this information is added to the caption). That mean it covers information about damage on both inner and outer ring but with the inner ring as the more damaged bearing ring as stated in the text.

- While reading this section, I was left wondering how severe the applied loads (1,000 kN axial for bearing 0411 and 210 kNm for bearing 0498) are compared to the static rating for the bearings or their static safety factors. That is, to what extent were these loads elevated and the test a highly accelerated life test?

  The 1,000 kN lead to a contact pressure of 2.6 GPa, the 210 kNm lead to 3.5 GPa. These load values and in turn relative high contact pressures were chosen in order to run the fatigue tests on the test bench at a reasonable time. This information will also be added to the paper as follows:

  The test for bearing 0411 applies 1,000 kN of pure axial compression force, **resulting in maximum contact pressure of 2.6 GPa,** and then rotates 720° back and forth with a maximum speed of 52°/s.

  The test for bearing 0498 applies a bending moment of 210 kNm, **resulting in maximum contact pressure of 3.5 GPa**, and pitches the bearing with a double amplitude of 80° at a maximum speed of 52°/s.

- Figure 8: I will admit that I didn't understand until re-reading that this Figure was not a progression like Figure 7, even though it was disassembled twice. Even knowing that, I am still having a difficult time interpreting it. I also can't necessarily understand the 80 and 280 degree locations between Figures 5 and 8.

  I admit that this figure might be challenging to understand. While Figure 7 only need to plot damage on two diagonals (because of the pure axial loading) it is easy to display, Figure 8 have to plot damage on 4 diagonals because of the bending moment and the load share due to the characteristic bearing behavior of a double-row four-point contact ball bearing. As the first disassembly with this bearing showed only very minor damage, the authors chose to just display the damage at the end of the test. This is also stated in the figure caption. The 80 and 280 degrees is just the position where the damage occurred and there is no direct link to the positions in Figure 5.

2.5 Bearing life calculation

- Lines 229 to 235: I had a hard time understanding what the authors meant in this paragraph. It took me a few readings to piece it together. I think the authors are trying to convey that the ISO 16281 method accounts for changes in load distribution around the bearing. Normally, for a pitch bearing, this would be due to the combination of radial and axial loads, a moment and the stiffness of the bearing. However, in the context of this paper, it also includes the lack of stiffness (or really, additional clearance) in areas of the raceway that were spalled and the resulting change in load distribution. If that understanding is correct, then I suggest that it be "These include the effect of the changes in load distribution due to the global flexibility of the bearing rings and the local additional clearance in the damaged areas of the raceways as shown in Figure 3. Hertzian contact is assumed wherever ball-raceway contact takes place – even in the damaged areas; the higher local pressures from non-Hertzian contact between the balls the damaged areas of the raceway is neglected. Furthermore, possible pressure increases due to truncation at the edges of damaged raceway areas is neglected. Therefore, the life calculation described in this paper aims to analyze the effect of changes in the bearing load distribution due to both localized damage and global flexibility."

  Thank you for contributing to a more understandable paper. The proposed changes will be applied to the text.

Results and Discussion

- Figures 5, 7/8, and 12 : Are the circumferential positions of all these figures the same? For example, is it a correct understanding that in Figure 12 the strain is low at the 90 and 270 degree positions because in Figure 7 this is the same position as the damage at segments 1 and 35 which is the same as positions 90 and 270 degrees in Figure 5? Also, I assume in Figure 12 that the 6 figures correspond to the 6 disassemblies (initial through 5). However, I'm not sure which is which. Can each one be labeled for clarity?

  Yes, the circumferential positions are all always the same. They are based on the blade bearing coordinate system according to DNV GL. The positive x-axis is always 0° and segment 1 for the figures 7 and 8 always start at 0°. Figure 12 names the different states in the legends of the plots starting with initial and going up to the 5th disassembly.

- Tables 2 and 3: I'm not sure I understand why most of the % deviations are 0 except for 1?

  Due to the scatter in the experimental data, the FE results are compared with the minimum and maximum values of the experiments. The aim for a successful validation of the FE results is to lay within this range or have a maximum deviation of 10% from the nearest experimental value for the largest strain. The deviation in

the table is 0% when the simulation data is within the range of the experimental data.

For more clarity the following statement is added to the text:

At all positions, the deviation between the simulation and the range of experimental data is far below 10%. **When the FE result is in between the range of the experiment, the deviation is listed as 0%.** The largest deviations can be seen for the 5th disassembly, indicating a limitation of the modelling approach of implementing damage in the FE bearing model when too much area of the raceways is damaged

- Figures 15 and 16: These are probably the most interesting figures in the paper, but I'll admit I'm not quite sure I understand them. Specifically, whether or not each point represents the estimated L10 for each load distribution (I think it does) or the accumulated L10. The text states "of an otherwise identical new bearing with the new load distributions", which I believe means it is the estimated L10 for each. However, the figures themselves feel like they are trying to communicate the accumulated fatigue life, or conversely, the remaining useful life. This question is relevant, I believe, because the introduction and a few other places in the paper discuss remaining useful life. I can't provide additional comment on the corresponding discussion and figures until I understand this point.

  These plots give an overview about the changes of the calculated bearing life L10 based on the load redistribution in the FE model due to considered damage. They do not cover any accumulated L10 or a true service life. Indeed, this is important. Therefore, the following will be added to the last statement:

  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.

  The authors understand this work as a basis for incorporating damage in FE simulations to detailed study changes in the bearing behavior during its service life. Accounting these changes in the calculation of the service life, however, is not part of this paper and will be topic in future research projects.

Minor comments:

Thank you very much for the minor comments. If not stated otherwise, they are all implemented in the text.

- Line 17. There is a mistaken period after "experimental."

- Line 29: “further” is used here, but further than what is undefined. I believe the authors mean “further than this first appearance of raceway damage (i.e. rating life)”.
- Line 31: Rather than “changes constantly”, I recommend simply “decreases” or possibly “decreases quickly” here. However, if the authors prefer to speak in terms of “rate of change of remaining useful life” rather than the “remaining useful life” itself, it is also possible with the current wording.
- Line 34: The acronym FEA is only use once, so there is no need to define it.
- Line 73: I think either “bearing service life” or “bearing remaining useful life” would be better here than just “bearing life”.

  This work introduces a methodology to consider raceway damage in FE bearing models and analyses changes in the load distribution and recalculates the bearing life to predict the likelihood of new damage initiation based on the altered bearing behavior

  If this is the mentioned part, we would like to keep it as bearing life, because in the end of this work we are not calculating the service life but just the bearing life based in the load redistribution.
- Line 90: The phrase “…any external load induces ball forces with both axial and radial components, leading to a tilting of the bearing rings…” I don’t believe is quite correct, as I don’t believe a purely axial load will result in tilting. Did the authors mean “…any external moment…” Or maybe “…often leading to a tilting of the bearing rings” would be sufficient?

  Actually, it does. For a four-point contact ball bearing with contact angles somewhere between 0° and 90° (usually around 40-45°) any external load is split into a radial and a axial component. Therefore, even a pure axial load introduces a radial force to the bearing rings. In addition, the two rows carry different amounts of the external load (especially on the traction side), leading in different radial forces on the two raceways and therefore resulting in tilting the bearing rings.
- Line 103: I believe the citation should be “in Graβman et al. (2023).”
- Line 290: Rather than “is not carrying any load”, shouldn’t it be “may not be carrying any load” or more precisely “is not carrying any load unless the bearing deformations are large enough to overcome the local clearance in the spalled area”?
- Line 335: I might be quibbling, but from Figures 12 and 13 I might recommend changing “successfully validated” to “reasonably validated”, but maybe this depends on my understanding of why Tables 2 and 3 have 0% in so many cases. I

believe it also depends on what I think of as “tailoring” of the negative clearance between 0 and 60 um in Figures 12 and 13, which one would not necessarily know prior to conducting the experiment.

The negative clearance of 60µm for that bearing is caused by a very tight fit between balls and raceways. As this is only visible in the experiment by a large friction torque and cannot be measured directly, this specific value is carried out iteratively for the FE simulations. As described in one of the earlier comments, 0% deviation means the FE has no deviation to the experiment. Because if this, the authors would like to keep the wording of “successfully validated” as it meets the defined criteria.