

The article compares the performance of two condition indicators derived from vibration measurements acquired during accelerated tests of Ø100 mm thrust bearings. The first condition indicator is a more traditional measure based on 2x the ball spin frequency from the Hilbert transform of the vibration, while the second is the z-score of the reconstruction error from a multivariate autoencoder of a number of similar condition indicators. The use of deviation of the reconstruction error from typical levels for a healthy bearing as a composite health indicator is interesting, and in general it performs better than the more traditional method in 4 of the 5 tests. I do like the article and recommend it for publication; however, I have a few general and several specific comments and questions for the authors to consider.

General

- It took quite some time for me to understand what is often described as 2 or more AI methods. This is true of the title (“AI enhanced fault indicators”, emphasis on plural), abstract (“two ways of using AI”), general descriptions in the article, and Conclusions (“two approaches”). I found this confusing and rather prefer the research question in Section 1.4 (“Can a combination of normal behavior model and high-level health indicator”, emphasis on combination). I recommend the authors make these descriptions more consistent. Really to me, I see what I described in my introduction to my review: the research value lies in using the z-score of the reconstruction error from a multivariate autoencoder of a number of vibration-based condition indicators. That is 1 method. At the end of the day, a single condition indicator (the “high-level indicator”) from the autoencoder is compared to a single traditional condition indicator (the “RPBF amplitude from HT+FFT”).
- With that understanding (if indeed it is agreed upon), I don’t believe the manuscript described exactly which traditional condition indicators were used in the autoencoder. That is, which formed ξ_i in Equation (9). I understand this may be considered proprietary, but if I could wave a magic wand this would be described better.
- In the Title and abstract much is made of AI. I will admit that I am not an expert in multivariate encoders (I did Google it), but my understanding of them is that rely on machine learning and/or neural networks. So, the “we must say AI” feels a bit much. Again with the understanding (if it is correct) that the research value lies in using the z-score of the reconstruction error from a multivariate autoencoder of a number of vibration-based condition indicators, I would just say that.
- The idea of verification is used twice in the abstract and once in the conclusions. I agree the bearing tests were quite nice, but fully “verifies” I think is a bit much. The “example results” in the Title is more honest.
- Combining several of these comments together the Title I think is more honestly something like “Autoencoder-based vs. classical vibration-based bearing condition

monitoring: example results of bearing tests”. Further, in the abstract lines 5 to 7 is more like “...less manual interaction. Because of the need for less supervision, this paper develops a single condition indicator from the reconstruction error of a number of vibration-based condition indicators using a machine learning-based multivariate autoencoder and compares its performance to a traditional single vibration-based condition indicator over 5 accelerated tests of Ø100 mm thrust bearings. In these bearing tests, the autoencoder-based monitoring...”

- Other than the mention of the need for testing on more complicated components like actual gearboxes, I’m not sure the “transferability to wind turbines” is really discussed in the manuscript. I’m happy to have the article submitted to Wind Energy Science; however, the meat of it is suitable for many publications such as MSSP, etc. Given that, I’d recommend “transferability to wind turbines” be removed from the title and “considerations of transfer to wind turbines” be removed from the abstract as they simply aren’t discussed. Wind turbines are just one of many applications of this method.

Introduction

- Line 25: My understanding of clause 7.6.1.3 in IEC 61400-1 is that it only allows a lower partial safety factor of $\gamma_n = 0.9$ for replaceable bearings in a structural ultimate limit state analysis. So, it just feels like this statement “IEC 61400-1 foresees reduced safety factors for monitored bearings” is a bit misleading. For example, the bearings discussed in the current article are most similar to bearings in the drivetrain (main bearing, gearbox, and/or generator), rather than part of the structure. Can the authors correct me if I’m wrong, or further explain what they mean? Another point to consider is that although IEC 61400-4 requires some bearings to be monitored, it does not include any special “credit” for doing so to my knowledge.
- Line 65: What does “monitor multiple indicators of different scales” mean in this context? In particular, what is meant by “scale”? I believe it refers to condition indicators with different units, for example an RMS versus a particular vibration amplitude (e.g. BPFO)? If this is the right sense, I might recommend simply “monitor indicators from different techniques”. The word techniques referring back to the 4 techniques described earlier in the section.

Results

- It is true that the indication is given earlier by the autoencoder-based method. However, in such an accelerated test it is difficult to understand how much earlier it is, especially since the x-axis does not begin at 0 in Figures 6 to 10. Maybe a more interesting metric to graph on the x-axis instead of test time would have been the proportion of L10

consumed? Further, a table summarizing the difference in time, possibly in %, across all 5 tests would also be helpful. I was flipping back and forth between the different figures several times in order to understand the difference in time, typically in fractions of a day except of course for the 2 tests where the traditional method gave no indication at all. Honestly, I'd focus on those 2 tests as being more significant than the 2 others where the auto-encoder was marginally better (and still be honest about the 1 test in which it wasn't).

Discussion and Conclusions

- Line 313: Here a change in bearing friction is mentioned as a cause of the change in bearing frequencies. I'm not sure this can be concluded, or even if it could why it matters? I'd recommend deleting this portion of the sentence.

Minor comments:

- Line 4 and 34: It's arguable the extent to which commercial systems "include" or "require" significant human supervision and interpretation of measurement data. I'd feel a bit better if this were softened a bit to "often include" or "often require", but I'm not gonna battle about it.
- Line 23: I think this sentence reads better if the latter half would be "...major repairs in wind turbines, continuous health monitoring of bearings facilitates more efficient maintenance thereby reducing operations and maintenance costs".
- Line 27: I believe "ISO28016-21" should be "ISO 20816-21".
- Line 41: CM is defined as the acronym for condition monitoring here; however, it is only infrequently (6 times) used in the remainder of the article whereas condition monitoring is used much more often (33 times). It is best practice to use the acronym consistently, or not at all.
- Line 42: The citation should be changed to "(ISO, 2015; ISO, 2016)".
- Line 55: A space is needed for "outer (BPFO)", while "and" is needed at "respectively, and ball spin frequency (BSF)".
- Table 1: A (?) is listed for a citation. I believe it should be Laouti et al., 2011. Additionally, DVSD is not defined elsewhere. Lastly, what is a "torque actuator"?
- Line 134 and 143: I recommend making these sentences part of the previous paragraph, rather than being their own paragraphs.
- Line 160: The BPFO, BPFI, BSF and I believe FTF were previously defined, so they do not need to be defined again here.
- Line 176: I believe alpha should be 90 here, not 0.
- Lines 183 to 191: Would be better as a single paragraph, rather than 3 paragraphs.

- Line 186: Would read better as “Example influencing factors include the background noise...”
- Line 227: Would read better as “The normal behaviour multivariate autoencoder trained on healthy...”
- Line 285: Hereafter RPBF is italicized, whereas before it is not. I don’t believe it needs to be.