ANOMALY-BASED FAULT DETECTION IN WIND TURBINE MAIN BEARINGS

Lorena CAMPOVERDE-VILELA, María del Cisne FEIJÓO, Yolanda VIDAL, José SAMPIETRO, and Christian TUTIVÉN

Response to reviewers

General comments of the authors

Dear Editor and the Reviewers,

We sincerely thank you for your constructive comments. Under the reviewers' comments and suggestions, the manuscript has been significantly strengthened both in contents and clarity. Below, you can see the changes that we made in response to each reviewer's comment.

The editor and reviewers found the paper of interest, yet they felt that several issues needed to be improved and clarified before the paper could be accepted for publication. In the revised manuscript:

- The changes made in response to Reviewer 1 are marked in blue.
- The changes made in response to Reviewer 2 are marked in red.
- The changes made in response to Reviewer 3 are marked in brown.

Reviewer 2

General comments

In my opinion, the publication represents a useful contribution to scientific progress in the context of WES. It is of interest to the entire wind power community. The main objective of the work is anomaly detection using simple Principal Component Analysis (PCA). The topic of artificial intelligence and machine learning is one of the hot topics of the moment. Therefore, it is also important to examine how these methods can lead to improvements in the context of wind turbines.

Author's reply: We appreciate your positive remarks on the contribution of the work to the scientific progress of wind energy systems and the importance of exploring the application of artificial intelligence and machine learning in the field of wind turbines.

It is important to the authors to show a simple solution that does not require additional sensors. The approach using PCA and the SCADA data is interesting here. However, I wonder whether, given the value of the turbines and the maintenance costs that may be necessary, additional sensors and higher-quality ML methods would not be more effective. The state of the art is that PCA is not particularly suitable for anomaly detection.

Author's reply: Thank you for this comment, for which we have improved the explanation of the real and practical utility of the proposed methodology in the revised manuscript. For this reason, we have added the following paragraph in the Introduction Section.

> Cost is a critical factor in the renewable energy industry, and wind turbines are no exception. While advanced sensors and machine learning methods can provide more accurate and comprehensive data on wind turbine health, they also come with a higher price tag. In contrast, the proposed approach aims to offer a more affordable solution that can be easily adopted by wind farms that lack condition monitoring systems. This approach may be particularly beneficial for older wind turbines that lack the built-in sensors and monitoring capabilities of newer models. By extending the operation of wind turbines close to their expected service lifetime, the proposed approach can help wind farms generate more electricity and revenue over time. This not only improves the profitability of the wind park, but also increases the overall efficiency of the renewable energy sector. The longer a wind turbine operates, the more energy it generates, and the more emissions it can help offset. Furthermore, the proposed approach could help reduce the environmental impact of the renewable energy industry. Manufacturing new wind turbines requires significant amounts of energy and resources, so extending the life of existing turbines can help to reduce the need for additional production, promoting a more sustainable and circular economy for wind energy.

Regarding the concern about the suitability of using Principal Component Analysis (PCA) for anomaly detection in wind turbines, we agree that this method may have some limitations, but it is well-suited for anomaly detection. In particular, PCA is a widely used technique for identifying patterns and trends in large data sets. By reducing the dimensionality of the data, PCA allows for the extraction of the most important information and the identification of the most significant factors contributing to the variance in the data. In the context of anomaly detection, this can help identify the most relevant features that contribute to the anomalous behavior. In our paper, the proposed approach using PCA (and SCADA data) has demonstrated promising results in detecting faults in the main bearings of wind turbines, as shown in the results obtained with real SCADA data.

Specific comments

Data preprocessing is not sufficiently described in the paper. If I understand it correctly, a range is specified for the real data and outliers are adjusted accordingly to the damage-free training data. This has several problems: Weak signals are filtered out, the model is only valid for the system under consideration, and the model assumes that the system under consideration is at the bottom of the bathtub curve, i.e. entirely error-free. Overall, due to the low sampling rate, the 10-minute intervals and the averaging over a week, the data appear to me to be very smoothed, which makes it difficult to find anomalies. Since we are dealing here with time series, a simple Pearson correlation is only of limited help (a Spearman's rank correlation should at least be examined here).

Author's reply: We apologize for the insufficient description of the data preprocessing in the initial submission. In the revised manuscript, we provide a more thorough and clear description of the data preprocessing. In particular, the following paragraph has been added, that also answers the reviewer comments about the problem of weak signals filtering out.

> In our study, extreme values (outliers) were not systematically removed since doing so could lead to a loss of information related to fault detection, as stated in Encalada et al. (2021). Instead, a strategy of defining ranges based on realistic values that can be obtained by different sensors was adopted. This approach, which allows potentially useful information to be retained while still addressing the issue of outliers, was chosen. To ensure appropriate definition of the ranges, non-restrictive criteria were used that were wide enough to encompass the majority of the observed data. By adopting this approach, it is almost ensured that the only outliers removed are those related to non-working sensors (not well calibrated or with faults) and/or due to problems with the communication of the data, rather than outliers related to the underlying physical process being monitored.

Thank you for raising the issue of the model's limited applicability beyond the specific system studied in our paper. The reviewer is correct that this is a potential limitation of the proposed approach. One way to address this limitation is through the use of transfer learning, which involves training a model on one dataset and then fine-tuning it on a new, related dataset. This can help to generalize the model to new datasets with different characteristics, and it is an area of active research in the field of machine learning. However, this is beyond the scope of this paper, as our goal was to develop and evaluate a model for each individual wind turbine based on its own data. We will consider exploring the use of transfer learning in future work, and this was added in the Conclusions Section with the following paragraph.

> Finally, while our approach has shown promising results, there are

several areas for future research. One limitation of our approach is its applicability to new datasets with different characteristics, as each WT depends on its own model trained with its own data. In future work, we plan to explore the use of transfer learning to overcome this limitation and develop models that can generalize to new datasets.

Regarding the comment about the correlation study, we agree that the Spearman's rank correlation is a useful tool for analyzing time series data, particularly when the relationship between variables may not be strictly linear. In response to the suggestion, we have re-examined our data using the Spearman's rank correlation, and we found that the results are consistent with our previous findings using the Pearson correlation. In the revised manuscript, the results obtained with the Spearman's rank correlation have been added, together with the addition of the following paragraph.

> Both Pearson and Spearman correlation coefficients are measures of the strength and direction of a linear relationship between two variables. The Pearson correlation coefficient is used when both variables are continuous and have a linear relationship. It measures the degree to which two variables are linearly related, and ranges from -1 (perfect negative correlation) to 1 (perfect positive correlation), with 0 indicating no correlation. The Pearson correlation coefficient assumes that the data are normally distributed. On the other hand, the Spearman correlation coefficient is preferred when the variables are not normally distributed or are ordinal (ranked). It measures the degree to which two variables are monotonically related, meaning that they move in the same direction but not necessarily at a constant rate. Like the Pearson coefficient, Spearman correlation coefficient ranges from -1 to 1, with 0 indicating that there is no correlation.

Finally, in regard to the comment about the low sampling rate, we agree that this issue was not clearly stated in the original version of the paper. The revised manuscript now includes the following added paragraph in the Results Section.

> It is significant that the proposed approach is designed specifically to detect (using only standard SCADA data, which are usually 10-minute averaged) the possible heat generated from an initial failure mode, such as the initiation or propagation of the crack, friction, electrical discharge and other failure modes associated with heat release. These types of failure typically result in a gradual and sustained increase in temperature (while they evolve), rather than sudden spikes or drops, which makes them detectable even with low sampling rates, as temperature variables have a low dynamic and

still contain the information of the fault after being 10-minute averaged. With respect to the use of the weekly average, it is intended to reduce false positives by smoothing out transient fluctuations in the data that are not indicative of actual anomalies. Although this averaging may limit the resolution of the approach, as it could smooth out subtle changes in the data that could be indicative of early-stage anomalies, this trade-off is necessary to minimize false alarms and ensure practical utility of the methodology (and avoid alarm fatigue).

The data show a clear seasonal component. The question arises why this was not removed by decomposition, especially since the model is only based on individual data sets. After peaks (Figure 11), the signal drops sharply again for longer periods. What is the reason for this? Since the level seems to be significantly lower in the period from May to November, it is questionable whether incipient damage could be detected here at all.

Author's reply: Regarding your question about the seasonal component in the data, we agree that this is an important consideration, and we did take steps to address it in our analysis. Specifically, we subtracted the ambient temperature to all variables related to temperature and used a rolling window approach to train our model on a subset of the data, which helped to capture the seasonality and other temporal patterns in the data. We acknowledge that there are other methods for removing seasonality from time series data, such as seasonal decomposition, and we will consider these approaches in our future work.

Finally, thank you very much for your comment regarding the after peaks signal dropping sharply again for longer periods. The reason for this has been explained in the revised manuscript, in the Results Section, where the following paragraph has been added.

> Note that after peaks (Figure 11, WT5), the signal drops sharply again for a long period. This is because the heat created from an initial failure mode (heating from an initial crack, friction, wear,...) is detected by the methodology, but its appearance is not continuous over time until the final breakdown. In contrast, when the failure mode advances, for example, when a crack propagates, the generated heat appears. When the crack remains still, no further heat is generated; thus, the alarm is set off. However, cracks are already present and can advance at any time, leading to the possible failure of the component. Thus, in this methodology, whenever the alarm is on (even when it is set off after a few weeks), it is highly recommended to check the specific WT.

Overall, despite the previously mentioned criticisms, the work is important because it helps to further advance the topic of machine learning and discuss the benefits and methods.

Author's reply: We are pleased to hear that you recognize the importance of our work in advancing the topic of machine learning for fault detection in wind turbines.

The anomalies also do not allow any statement on the type of damage present and the severity of the error. At the same time, there is no statement about the historical data and any false alarms. For a scientific consideration, a hit rate AND an error rate must be given.

Author's reply: We acknowledge the issue regarding the capability of the model in detecting the type of damage present and the severity of the error. We believe that further developments could be pursued in this direction, for instance, by incorporating high-sampling rate data and/or additional sensors to improve the precision of the fault location. However, we also recognize that this may come at the cost of increased complexity and expense, that we are trying to avoid in our contribution. While our method may not provide detailed information on the type of damage or its severity, it can still provide valuable insights into the system's performance and indicate the need for further investigation or maintenance actions. As not being able to adequately locating the fault is a clear limitation of the proposed methodology, in the revised manuscript we added the following paragraph in the Conclusions Section (highlighted in blue color as Reviewer 1 also commented on this issue).

> While the main bearing temperature was found to be a suitable indicator for detecting faults in wind turbines, as also stated in a recent paper by Murgia et al. (2023), another limitation of the proposed approach is that it cannot precisely locate the fault or its severity. Further developments could be pursued in this direction, for instance, by incorporating high-sampling data and/or additional sensors to improve the precision of the fault location. However, this may come at the cost of increased complexity and expense, which is trying to be avoided in this work where the main objective is to contribute a cost-effective solution where all variables used are readily available in all industrial-size wind farms (both older and newer).

In regard to the proposal to incorporate a hit and error rate, we thank the reviewer for taking this into our attention. The revised manuscript incorporates the following paragraph in the Results Section.

> In summary, 18 wind turbines were examined, of which 16 were considered healthy and correctly classified as such. One turbine

had the fault of interest and was correctly classified as faulty. Another turbine had a fault (that was not the fault of interest) and was classified as faulty, which could be considered a false alarm. However, in practice, the fact that an alarm was raised for a fault in a different component could still be useful, as it indicates the need for maintenance or further inspection. Therefore, in addition to the hit rate and error rate, the practical implications of false alarms should also be taken into account.

The metrics in Table 4 need to be explained.

Author's reply: Thanks for bringing this issue to our attention. The metrics in Table 4 are now detailed in the revised manuscript.

Line 133 and Figure 3: It is not described what kind of damage is typically involved here (lack of lubrication, wear, pitting, ...). Each defect should produce certain characteristics in the measured variables.

Author's reply: We agree that each type of damage could produce unique characteristics in the measured variables. However, in our case, the only information available regarding the fault is the work order information stating "Replacing Main Bearing." This limited information makes it difficult to determine the exact type of damage involved.

The structural, linguistic and graphic quality of the publication is very good. The work is clearly structured and the tables, graphs and pictures are easily recognizable and informative.

Author's reply: We believe that clear presentation and effective communication are essential in scientific publications, and we are delighted that our work meets these standards. We appreciate your review and will continue to strive for high-quality presentation and communication in our future work.

Technical corrections

Figure 16 with WT16 should be placed closer to line 326 where it is addressed.

Author's reply: We appreciate your attention to detail and your effort to provide constructive feedback that can improve the readability and clarity of our work. We placed the figure closer to line 326 in the revised manuscript.

Finally, we would like to thank the reviewer for the valuable feedback and the time to review the paper.