Review: Simulating run-to-failure SCADA time series to enhance wind turbine fault detection and prognosis

This work addresses the failure detection and prognosis in the context of wind turbine operation. The paper introduces a synthetic data generation methodology for the training of failure detection and remaining useful life (RUL) prediction algorithms by using cGAN to generate SCADA data abiding to predefined conditions. The methodology aims to improve the prediction accuracies of the algorithms by providing more data samples that can better represent degradation trends of the wind turbine. SCADA dataset from a wind farm was used to validate the methodology.

In my opinion, the manuscript addresses an important research topic that is very relevant and within the scope of WES journal. The manuscript provided a sound methodology, and the content is a valuable contribution to the research area discussed.

Overall, the language of the paper is well written tonally, and the figures were clear and helped in presenting the findings. The manuscript provided a very good overview of the problem statement, while also providing relevant literature review to address the limitations of past works. The methodology is mostly well discussed and presented, but the structure of the result sections can be improved to provide more clarity for the steps taken to reach the conclusion. The conclusion is concise and summarised the findings well, but the limitations of the work can be elaborated.

My comments and questions will be listed below, points that in my opinion need clarification will be listed in *Remarks* and editorial suggestions will be listed in *minor comments*.

Remarks:

1. pg.4 - “the weights of the fR and fstd terms are experimentally set to 3 to balance the four terms”, what do you mean by “experimentally set”? Was it by iterations with a sample data-point?
2. Pg.9 – Section 3: Dataset, what is the test-train split strategy adopted? Why is only WT9 data used as validation set instead of sampling from all the wind turbines?
3. pg.10 – “SMOTEN method”, I believe this is a typo. Which dataset was resampled? What value was resampled and why?
4. Pg.12 – “The Adam optimiser ….”, it appears that hyperparameter fine-tuning was performed to optimise the algorithm, it will be good to include the fine-tuning strategy adopted to reach this conclusion.
5. Pg.12-13 – Results on misclassified labels and false positives, it is interesting to see that despite having misclassified labels, most of the WTs had 0 false positives. Would you not consider this to be a sign of overfitting?

6. Section 5. RUL prediction case study, what is the baseline and structure of your performance measurement for the proposed method? Discussion on why the monotonicity value from the MK metric is relevant to the quality of degradation trend and how it can affect the RUL prediction should be included.
7. Pg.16 – Justification to why the second-order polynomial function is used to predict RUL is needed. The method is only tested with WT6 dataset, this makes me wonder if the same method will be applicable and effective on a different failure case from a different wind turbine.

Minor comments (suggestions):

1. In pg.1 line 13, "… to produce reliable RUL ~~estimates~~ estimations."
2. In pg.1 line 24, "… improving their robustness and ~~practical applicability~~ practicality."
3. In pg.10, line 200, "the effectiveness of the developed method in fault detection." Can be clearer on which developed method this is referring to (e. of the developed synthetic data generation method).
4. For section 4., a short separation sentence can be included to clarify that the result of the HI produced from (a) SMOTE generated data and (b) cGAN generated data will be compared and discussed. I also find that separating [texts addressing model configurations and methods adopted] from [result presentation] into different paragraphs can improve the structure of the section.