Rebuttals

Based on the action editor's recommendation, your article needs to be revised and reconsidered. For your guidance, reviewers' comments are appended below.

If you decide to revise the work, please submit a list of changes or a rebuttal against each point which is being raised when you submit the revised manuscript. The Guide for Authors can be found via the URL below.

Thanks for the constructive comments. We have done a major revision of this manuscript by considering all reviewers' comments. All the changes have been highlighted in the revised paper. In addition to the response to the reviewers' comments, we have revised the entire manuscript to improve its readability, highlighted the main contributions of the work and provided more and deeper explanations of the research outcome. The result is a much improved manuscript. We hope that it will meet your approval.

Reviewer 1's Comment:

Reviewer #1: The paper proposes a general framework for Remaining life estimation of mechanical systems applied to a NASA benchmark dataset. It can be can be considered for publication provided that the authors can address the following issues:

Thank you for your constructive suggestions and affirmation of our work.

1) One of the main claims of the authors is that the proposed general framework is computationally efficient, compared to other methods. This is not obvious at all since the proposed method combines several model selection strategies including evolutionary algorithms, and an initial evaluation on 6 models. So, there are lots of calculations involved to get the final model. More discussion is needed.

Sorry for the confusion, we made sure that the current revision of the paper is clearer with regards to this point. We mean that the resulting model chosen by the proposed framework is computationally efficient, particularly when it is applied to large sets of data in real applications. The framework does take some off-line computations to obtain the optimal model.

2) How did the author determine the redundancy of the information when selecting the optimal sensor readings?

The sensor readings are selected based on previous studies, which are contained in the references of the paper. In short, some of the sensors do not vary at all throughout the engine's life cycle while others are redundant according to PCA. This discussion is provided in Section 3.2.

3) The authors mentioned "Re" is randomly selected from 95 to 140, but in the next page they select [90,140] as bounds of optimization. Please correct this if needed.

It was indeed a typo. It is corrected in the revision.

4) The discussion in section 4 needs to be extended, and section 3 can be curtailed a bit.

We extended Section 4. In brief we included new statistics, new plots and expanded the discussion of the results. Section 3 was extended a little bit as well in order to address some of the concerns of the other reviewer.

Overall, this is a nice research and have some merits, provided that the authors be able to address the raised issues.

We hope that this revision meets your expectations and clearly addresses your questions and concerns. Thank you.

Reviewer 2's Comment:

A key aspect in the field of Prognostics and Health Management (PHM) is to avoid failures of mechanical systems. The accurate estimation of remaining useful life (RUL) is one way to help determine when such systems or components may fail in the future; thereby providing much needed lead time for corrective action(s) to be taken.

The authors present their method and apply it to a publicly available dataset of simulated aircraft engine data created by the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) at NASA. The public data was used as challenge data for PHM'08 and contains a subset of 21 features (called sensors) out of a possible 58 that are available from the simulator; with associated challengers being ranked and the best performers reported at that time.

Major comments
==========

1. The paper reports performance statistics/scores (RMSE, etc) of their results, which is good. However, the comparison of results in Table 10 does not include all results from the literature. For example, the authors reference [16] has a reported result of 7.02. There are, of course,

qualifications about that particular result, but the readers should learn of it within the author's paper so that a reader knows that the reported result of 14.39 (which may or may not be the best) may have room for even further improvement and hence that it may be a hard problem for the scientific community to sufficiently address. In addition, it is not clear which results were reported at the time of the PHM'08 competition; that information should be clearly marked in Table 10.

We have already addressed this issue. Although it is impossible to include each and every result reported on the C-MAPSS dataset we included some other results in Table 10. The criteria for including a method is that it is machine learning related and that it is recent. Further details are provided in Section 4.3.

A comparison against the results reported at the time of the competition are hard to obtain since very few of them were published. Furthermore, the presented scores correspond to a special subset (FD005), which was only used at the PHM08 competition and is not available to the public. Therefore, such comparisons are not possible.

2. The paper does not clearly present the RUL results because the paper focuses on the performance of the RUL results. It would be very helpful for the paper if the RUL results themselves were clearly presented and/or explained. In particular, the authors should ensure that they clearly explain that the competition had training and testing data and that the validation data was held back and only published after the competition. The authors should convince the reader that they have not used the validation data to help them gain an unfair advantage over those competitors. I have just seen figure 2; which does help the paper but it isn't clear if this is the best result of the paper or is an example.

We have included a new Section 4.2 called Experimental Results, we present the results with some plots and tables for the reader. We explain that we did not use the validation data to train or tune our method in Section 4.

3. The authors should clearly define what they mean by: shape, shaping the data, quality, efficiency and reliability. Figure 1 is not clear when printed and is also not clear conceptually for a reader... perhaps addition of 2 examples in this figure may help.

We made the definitions clearer. Figure 1 is now complemented with another figure plus an example of how the moving time window is generated for the C-MAPSS dataset. Regarding the quality of the figures, we have improved them and they are now included in the paper. If needed we would be happy to provide the original high-quality figures in a separate cover.

It is not entirely clear what neural network architecture is being used by the authors. They mention MLP and when Appendix A tables A.1 to A.6 are inspected it is not clear how many layers, how many neurons in each layer, and which aggregation function.

> *The tables in the appendix describe how a neural network model is formed. The following description was added in Section 3.1 and the appendix in order to make the tables more understandable.*
>
> *"Each row in the table represents a neural network layer while each column describes each one of the key parameters of the layer such as the type of layer, number of neurons in the layer, activation function of the layer and whether regularization is used, where L1 denotes the L1 regularization factor and L2 denotes the L2 regularization factor."*

I presume that "Activation" means activation function and if so, the definition of ReLU and Linear should be clearly presented in the paper (please confirm).

> *We created a new appendix where we include the definitions of the used activation functions and a simple plot of them for better readability.*

Also, the authors need to clearly state how many neurons are in the input layer and why they might be different for each neural network being trained. I can guess about this, but as a reader, I shouldn't have to guess because the paper should tell me explicitly. In general, the paper should always favor being explicit over being implicit.

> *The number of input neurons is now clearly stated in Table 8. Furthermore, a discussion of how the number of neurons may change depending on the selected dataset is included in Section 3.2*

As a side note, if it is possible to clearify in the paper when 20 epochs versus 200 epochs were being used that would help; I did get a little lost on that point... I am suspecting that the paper would benefit from a section called Neural Network implementation where the authors clearly explain how they implemented what they have done and therefore the section currently called "The Network Architecture" can be used to clarify the neural network concepts.

The paper would also benefit from the addition of a new section called "Experimental Settings". The authors may then explain what experiments were performed and their associated settings... leading to the results of the paper being presented in a section called "Experimental Results" and finally a section called "Discussion" (or, as the authors currently have it, a section called "Evaluation of the Proposed Method").

We followed the advice and split Section 4 into three subsections. Section 4.1 presents the experimental setting. Section 4.2 presents the results and a discussion of them. Section 4.3 compares the results obtained by the presented method with some state-of-the-art methods.

4. The authors have published their source code for the paper publicly; which is a very kind thing to do. However, they have not put their source code URL into the references section; they need to do that because all references should be placed into the references section when writing a paper. It would also be helpful for the community (not for this paper submission) if the authors added a README in the following location to help orient a person trying to figure out what are the high level source code files and which are supplementary: https://github.com/dlaredo/NASA_RUL_-CMAPS-/tree/master/code

We included the project's URL in the references section.

5. When looking at the paper from a high level, it can be seen that approximately 8/10 pages are written explaining background and 2/10 pages are written explaining results. In order to help balance and improve the paper, it would be good if 2-4 or more pages of results were carefully explained and presented for the reader. In other words, experimental evidence is what readers of a scientific paper are looking for... so add as much as is necessary to clearly support the paper's reasoning and conclusions. For example, "it is common to observe that the parameters leading to lower score values in the early stages of training are more likely to provide better performance...". It would be great to support this comment with evidence; a figure, a statistical result, etc. Perhaps comments like this would require an experiment to confirm or deny them... please do what is needed.

We extended the discussion on the obtained results. The extension also includes some plots and Tables to help illustrate the concepts and ideas discussed. Section 4.2 was created to addresses this issue.

Minor comments
==========

1. CMAPS is used in the paper and not defined; it should be defined. Perhaps the authors are referring to C-MAPSS, the publicly available data?

2. Please avoid using the word "recall" in the paper because it is informal and if the paper is incorrect, it becomes impossible for a reader to recall something that may not have been written at all. It would be better to refer to a section, equation, table, figure, etc. explicitly because it helps the writer ensure that the information exists and helps the reader find the information to recall.

3. Please ensure the paper uses "implies" correctly... the paper may mean "requires" instead.

4. Algorithm 1 should have all values clearly defined and those that are inputs should be in the inputs section. For example, the number of epochs seems to be an input to the paper's algorithm. What is n? Is it related to X? Is y a vector of values?

Each one of the issues mentioned here has been taken care of. We thank the reviewer for the detailed list of corrections, which have helped to improve the quality of the manuscript