The article, titled " Overview of normal behavior modeling approaches for SCADA-based wind turbine condition monitoring demonstrated ondata from operational wind farms" covers a very interesting topic that perfectly fits in the journal's scope.

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.

However, there are a number of major issues that need to be amended or clarified.

Major issues

2.1 Preprocessing techniques.

A list of different techniques for filling/treating missing values is given, but it is needed to give a comprehensive explanation of the benefits/advantages and disadvantages of each one of them. When/why is it recommended to use one or another?

Related to outliers removing, it must be explained how to avoid that abnormal values associated with the failure of interest are also removed.

3.2.2 Selecting healthy training data

It is not clear that healthy data coming from data previous to a failure can be used together with healthy data coming from a data posterior to a failure to construct a NBM. That is because the replacement of a component can significantly affect the (normal) behavior of the WT. Some authors even recommend a "quarantine" of data not to be used after the replacement of a major component such as a gearbox. The WT with the "new" gearbox can have different normal behavior, as in fact it is a different machine. In the paper figures 4 and 5 are too simplistic and do not take the aforementioned comment into account. Explain, how this can be considered or counteracted.

3.3 Normal behavior modelling
In this section, it is written that "The explicit NBM that will mainly be used is the elastic net. It is a simple, transparent, and robust model that can handle large amounts of (correlated) predictors, while at the same time it can work with a limited amount of training data. This corresponds to requirements set by the industry. For this reason, deep learning models, like AEs, are not used in this research."

It is needed more detail about the statement that "industry requires a limited amount of training data". Why this requirement? How limited? The reasoning that deep learning models are not used in this research because of this "requirement of industry" its weak. It is missing an important comparison, as AEs are unsupervised by nature and one of the best fit models for this type of problem. Testing only with elastic net is too simplistic, as deep learning approaches are not compared (only SVR and lightGBM wchich are machine learning approaches). I recommend to compare at least with ANN.

3.4 The anomaly detection procedure

Explain the meaning of variable idio_comp in eq. 5.

Explain the meaning of q in eq. 6.

It should be detailed how variable health_score is computed.

Please, review all equations and explain in detail all variables involved, as they are not clear overall the manuscript.

4.2 Experiment 2: The added value of using lagged predictors
It is said “Reasons for the low added value of the lags can perhaps be an insufficient number of lags, a lack of information on the dynamics in the aggregated SCADA data, or the combination of transient and non-transient behavior. The first hypothesis seems to be unlikely since limited experimentation using more lags showed no clear improvement in performance. The second hypothesis is possible. The third hypothesis would imply that the dynamics of the steady-state and the transient behavior of the turbine are so different they can not be learned by one elastic net model.”

I do not agree with the statement. Temperature variables have a slow dynamic, and thus the second hypothesis is not likely to be correct. However, the third hypothesis is very likely, as the WT behavior is highly non-linear, and thus can not be learned by one elastic net model. This is the reason lags did not show improvement. Again, testing only with elastic net is too simplistic, as deep learning approaches are not compared that can really “learn” to highly non-linear dynamics that the WT undergoes.

The experiments given in Sections 4.1, 4.2, 4.3, 4.4, 4.5 are not really tested for early failure prediction, but only tested for how well the NBM predicts the target variable. The title of the paper states that it is an “overview of NBM approaches for SCADA-based wind turbine condition monitoring”, therefore this implies testing not only the NBM itself but also the analysis of the NBM prediction error. It is too simplistic to only compare the prediction error on the target variable. It is necessary to include in sections 4.1, 4.2, 4.3, 4.4, 4.5 a study about the prediction error and how many false alarms and good predictions of faults are obtained. For a scientific consideration, a hit rate and an error rate must be given.

In summary, the review poses many questions/hypothesis but fails to focus and answer, through in depth analysis, many of them that are vital. For example:

- Which is the impact of each one of the preprocessing techniques listed (on the early fault detection objective)?
- Which are the benefits/advantages and disadvantages of the different listed techniques for filling/treating missing values?

- Related to outliers removing, how to avoid that abnormal values associated with the failure of interest are also removed?
- Which are the best AI techniques to be used? Here a drawback of the paper is that it is avoiding testing a very important representative part of them (the ones based on deep learning).
- An analysis about how different algorithms for the analysis of the prediction errors affect the final results (number of false alarms and correct alarms on the 5 wind farms over the testing period must be given).

Minor

Reference Catellani et al., 2021 is incorrectly spelled, it should be Ca**s**tellani.