**Answers to reviewer 1**

We would like to thank the referee for the remarks. In what follows, an answer is given to each remark.

*The literature review is reasonable but not complete. Useful research has been published on the application of Copula models and also Gaussian Process models in wind turbine normal behaviour modeling. These should be included in the review.*

> The authors of the manuscript acknowledge the fact that research using copula and gaussian process models was missing. The literature overview has been expanded. Copula and Gaussian process based research has been added. The authors would however like to point out that the scope of the literature review has been limited to research that uses normal behavior modelling on temperature signals, meaning that the target should also be a component temperature. A lot of research that uses Copulas and Gaussian processes focuses on the power curve, where the target variable is the active power. This means that this research falls outside the scope of the literature review. The reasoning behind limiting the scope of the literature review is the fact that the topic of the paper focuses strictly on modelling component temperatures, but also that expanding the literature review to also contain the power curve research would inevitably result in a loss of depth of the literature review.

*I'm not clear why the data filtering reduced the normall 10 minute data to hourly. Better results may have been obtained using the 10 minute data directly (after error and gap removal). If noise reduction was the main reason this should be demonstrated statistically.*

> The reasoning behind reducing the resolution of the data to hour level (from 10 minute level) through aggregation is threefold. Firstly, component temperatures tend to change relatively slowly over time, which makes them more suitable for the detection of relatively long-term evolutions. This means that for this case the difference between the 10-minute or 1-hour resolution is marginal or non-existent, something that became clear during experimentation. Secondly, aggregating the data results in noise reduction. This was shown in Turnbull (2021). Thirdly, it reduces the computational burden of the methodology considerably, which is one of the main focus points of this manuscript. The authors acknowledge nevertheless that there are cases where this trade-off should not be done. This is the case for failures that form very fast over time, or for signals that exhibit damage patterns that are very short-lived like in vibration analysis. This explanation has been added to the paper.

*The mathematics of the modelling and error prediction should be presented in more detail. The terms in equation 2 should be explained.*

> The authors of the manuscript acknowledge the fact that the equations were not sufficiently explained. A more thorough explanation of the equations has been added to the revised manuscript.

*Cooling transients are a major cause of modelling difficulty. However they do not represent an operational turbine and I'm surprised such data was not excluded from the analysis by using status or power data from the SCADA,*

The power downs of the turbine have been removed from the data using the active power as a guideline. The transient behavior has also been removed to a large extent from the data by subtracting the fleet median from the signals. Figure 2 shows that transient behavior that is common to the fleet is captured by the common component (fleet median), and that the idiosyncratic component is free from it. This type of transient behavior is caused for example by cool-downs due to lack of wind, … What remains of transient behavior is unique to the turbine. This can for example be caused by maintenance of the wind turbine. These events are however much more rare. So most transient behavior has been removed by subtracting the fleet median from the signals. Furthermore, the data has been aggregated to hour-level, which further reduces the remaining transient behavior.