Referee#2: Please find below the answers to the individual remarks ordered from general to specific.

Q1: In the manuscript, it is mentioned that data 33 wind turbines from 3 different sites are analysed in this paper. Do they represent the same turbine typology, i.e. geared versus direct drive, synchronous generator versus DFIG etc.? What kind of site conditions they represent, complex terrain versus flat terrains? Age of the wind turbines? It is important in my opinion to discuss the representativeness of the data that are used in the analysis as the Resulting method will be applied to different turbine types, ages, site conditions etc... .

A1: The turbines are from different manufacturers and all of them are geared and equipped with DFIGs. All turbines were commissioned later than 2013 and the analysed periods fall within the first five years of operation. The sites can be characterised as moderately complex with mild elevation changes and occasional vegetation. This information was indeed missing and will be incorporated in section 2.1 of the revised manuscript accordingly (revised manuscript: p. 4, line 14 ff.).

Even though the turbines represent a rather homogeneous set we expect the method to perform equally well on temperature measurements along the drive train from turbines with different configurations where different sensors might be in place. This is due to the method's good performance over the wide range of different temperature signals as well as the different characteristics of the detected change-points. Neither do we expect the method's performance to decrease for older turbines or turbines in different site conditions. However, these characteristics might influence the presented cp statistics, with older turbines or turbines exposed to higher loads showing an increased amount of change-points due to increased wear and consecutive maintenance actions. This line of thought was added to section 2 as well: 'Even though these findings might vary across different turbine types, ages and site conditions the order of magnitude of CP presence highlights the necessity of a robust CPD methodology.' (compare revised manuscript p. 6, line 14 ff.).

Q2: The data analysed here are temperature data according to the table B.1 What is the reason behind this choice?

A2: In SCADA based monitoring of wind turbines using NBMs two approaches can be distinguished - performance and temperature monitoring. The former aims to detect abnormal deviations from the turbines usual power output, whereas the latter aims to detect deviations from the healthy thermal equilibrium conditions. Although both approaches have proven to be valuable (particularly in combination) temperature monitoring is better suited for detecting malfunctions in the components along the drive train, which account for the majority of turbine downtime (compare [1]). Moreover, the challenge of change-points in wind turbine SCADA data was mainly reported in the context of temperature monitoring in literature. Therefore, we decided to focus on temperature data. Nevertheless, the methods performance over a wide range of different temperature signals as well as over the different characteristics of the detected change-points suggests that the method can potentially be extended to other signals found in SCADA systems, a proposition that has been incorporated into the outlook section 6 of the revised manuscript (compare revised manuscript p.21, lines 11 ff.). Thanks to the referee's comment it also became clear, that neither the distinction between temperature and performance monitoring nor our motivation for variable selection were stated explicitly enough. Therefore, they incorporated into the introductory section (compare updated manuscript p. 2, line 21) as well as the data set description of section 2 (compare revised manuscript p. 4, 18ff.).

Q3: Are there vibration data from the wind turbine also available for the analysis. Will the algorithm change if other types of sensors are analysed, e.g. acceleration data?

A3: Vibration/Acceleration data were not available for this study. Nevertheless, we assume that in principle the suggested kernel-based change-point detection algorithm should also be useful to analyse measurements from these kinds of sensors. [2] for example presents experimental results of kernel-based change-point detection being successfully applied to the segmentation of audio signals. In terms of structure and time resolution, audio signals are much closer to vibration data than the SCADA data analysed in this study. One particular challenge we see at this point is that the high data resolution could impose numerical challenges for computing the respective gram-matrix. In any case, the proposed data pre-processing method would need to be adjusted to the different types of data and could potentially help to overcome these problems. We think this is an interesting question that could be addressed in the future and therefore incorporated it into the outlook section 6 (compare revised manuscript p. 21, 10 ff.).

Q4: Change of operation modes. Does the algorithm consider changes in the operational state of the wind turbines? For example, downregulation of power due to grid demand, noise-reduced operation due to noise regulation in the night with medium/high windspeeds. These can look like CPs in the data possibly.

A4: The proposed framework considers changes in operational states of the wind turbine in two different ways. Firstly, the pre-processing procedure acts as a normalization which puts the measured temperature in relation to the operational state. Secondly, by averaging the signals over a full day, which was originally motivated by computational considerations, the impact of such presumable sub-day events is further reduced. This helps the algorithm to focus on the most significant and long-lasting changes and is part of the reason, why the pre-processing has such a crucial effect on the algorithm's performance (compare section 5.3).

Q5: Minor comments and edits can be found in the attached PDF file:

i. **Q5.1:** The pre-processing takes care of seasonal effects. What about diurnal effects?

   A5.1: To reduce the numerical effort of computing the gram-matrix a daily averaging of the signals is part of the pre-processing procedure. This also removes all diurnal effects. Moreover, diurnal effects would be detected only with a penalty much lower than the proposed one, since the reduction in the cost function would need to compensate for as many change-points as days in the analysed period. Seasonality on the other hand induces 2 to 4 false CPs in each seasonal signal when not handled prior to the CP optimisation step and is therefore much more likely to be flagged by the algorithm.

ii. **Q5.2:** Page 4, line 4: 1Hz sampling is usually possible, the only problem is they are not being stored due to data storage reasons. The second reason you don't see them is because OEMs don't give access to wind farm operators.

A5.2: We agree with the reviewer and have updated the manuscript accordingly (compare revised manuscript p. 4, lines 3 ff.).

iii. **Q5.3**: Notes on spelling/grammar in the PDF-file.

A5.3: We agree with the reviewer and have updated the manuscript accordingly.

REFERENCES:

[1] Dao, C., Kazemtabrizi, B., Crabtree, C.: Wind turbine reliability data review and impacts on levelised cost of energy, Wind Energy, 22, 1848-1871, 2019

[2] Arlot, S., Celisse, A., and Harchaoui, Z.: Kernel change-point detection, arXiv preprint arXiv:1202.3878 ,2012.