

Julia Walgern

March 13<sup>th</sup>, 2026

**Response to Decision on submission to Wind Energy Science for manuscript number wes-2025-212 entitled "Reliability and O&M key performance indicators of onshore and offshore wind turbines based on field-data analysis"**

Dear Editor, dear Reviewers,

Thank you for the decision letter regarding our manuscript. We are pleased that our manuscript is being considered for publication and we would like to thank the reviewers for their valuable comments and suggestions for improvements. It is believed that the quality of the paper has been further enhanced by incorporating the attached list of modifications in response to the reviewers' comments.

In order to facilitate the identification of the changes made, we have made modifications in track-change mode. In addition, a "clean-editable" copy where the changes are not marked is attached.

Kind regards,

Julia Walgern

## **Reviewers' comments including answers by the authors (in colour):**

### **Reviewer 1**

1. Normalization of Failure Rates (Per Turbine vs. Per MW): The authors state that onshore wind turbines have higher failure rates per megawatt than offshore wind turbines, while they have lower failure rates per turbine. Even though this is an interesting observation, the decision to focus the remainder of the analysis and conclusions on the "per MW" metric is significant.

From an O&M logistics perspective (vessel transfers, crew planning), a "visit" is a discrete event regardless of the turbine's capacity. A 10 MW turbine failing does not necessarily require twice the number of visits as a 5 MW turbine, though the cost of the spare part or the revenue loss might scale. However, throughout the manuscript, the authors either implicitly or explicitly assume a linear scaling of failure rates versus turbine size. As an example, in line 368: "As detailed in Section 5.3, onshore WTs experience an average of 1.088 failures per MW per year. For example, this is equivalent to 5.4 failures per year for a 5 MW turbine and 10.9 failures per year for a 10 MW turbine."

The authors should elaborate on the justification for assuming a linear scaling of failure rates with capacity. Does a control system or a sensor really fail more frequently just because it is installed on a larger machine? It is recommended to include a brief discussion on which subsystems scale linearly with power (e.g., perhaps converters) and which do not, to prevent readers from misinterpreting the "per MW" metric as a universal rule for reliability scaling.

We fully appreciate the point of view of the reviewer here. In Section 4.1, we therefore highlighted already "However, it is important to note that the failure behaviour is not solely characterised by turbine size making more sophisticated reliability models necessary to support such analysis."

While "~~similar~~ intervention frequencies ~~similar to these~~ are observed across offshore wind farms with different WT power classes included in the data sets used for this analysis" (Section 4.4.1), it is true that not all subsystems of the turbine will scale with rated power, and those that do are unlikely to do so linearly. This is something we investigate within another manuscript, which has just been accepted for publication (see response below). We have added the following discussion in accordance with the recommendation of the reviewer to Section 4.4.1:

*"Readers should note that this implicitly assumes a linear scaling per MW for all subsystems, which is a simplification. We have explored this in more detail in (Walgern et al., 2026; Walgern et al., 2025), and for the converter system in (Anderson et al., 2024; Anderson et al., 2025). In that more detailed analysis, the control system, yaw system, lifting gears and meteorological measurement system show no evident trend with regard to rated power. The rotor, pitch, hydraulic, power generation, generator switching and converter systems do. For more accurate insights into scaling for these subsystems, readers are referred to (Walgern et al., 2026; Walgern et al., 2025)."*

Given the large amount of data available, a quantitative analysis of this trend seems important. This analysis must be performed separately for onshore and offshore turbines to isolate the effect of operation environment and focus on the effect of size.

Again, we fully agree with the reviewer's perspective that a quantitative analysis of this trend is important. However, we consider it outside the scope of this analysis for the simple reason that

there are several confounding factors on top of rated power and whether the turbine is onshore/offshore. There is also the turbine commissioning date (as we have observed this also has an influence), the design choices of the turbine, and the widely varying observation periods (due to censoring) that make up our field dataset. The purpose of the above-cited paper (Walgern et al., 2026) is to separate and quantify these coinciding effects.

2. The "Short" Useful Life Phase: The analysis of failure behaviour over time identifies a bathtub curve. However, the authors note that the constant failure rate phase ("useful life") is "surprisingly short," lasting only from years 5 to 8 before degradation sets in. This is a critical finding for life extension and financial modelling. Could the authors provide more hypothesis or insight into why the degradation phase begins so early (Year 9)? Is this driven by specific components (e.g., pitch hydraulics or blade erosion) that have shorter lifecycles than the 20-year design life? Highlighting the specific subsystems driving this early rise in the wear-out phase (as hinted at in Figure 5) in the text would strengthen the practical implications of the paper.

Again, we appreciate the perspective of the reviewer, and this is something we want to explore further in future publications. At present however, we unfortunately can go not much further than identifying trends in failure behaviour at the current subsystem level. Again, we have explored failure trends w.r.t time in more detail in our other paper (Walgern et al., 2026).

For reference we have added the following sentence to Section 4.3:

*"A more detailed analysis of failure trends regarding time can be found in (Walgern et al., 2026)."*

3. Minor comments:

1. Figures 3 and 4 use colour coding for OEMs. Ensure these colours are distinguishable in black-and-white prints or for colour-blind readers.

We thank the reviewer for raising this point. In Figures 3 and 4 the colours are used only to differentiate OEM groups visually. The analysis does not rely on the specific OEM identity; rather, the figures illustrate the spread in failure rates across groups. As the journal publishes figures in colour and the interpretation does not depend on distinguishing colours in black-and-white, we retained the colour scheme.

2. Line 276 "Given the strong dependence of average failure rates on WT size": This sentence is a bit ambiguous. It is better to explicitly state what kind of dependence is meant (larger turbines fail more often).

The wording has been changed in accordance with the reviewer's wishes:

"Given the clear trend of higher failure rates for larger turbines"

## **Reviewer 2**

This paper provides recent, detailed field data on the reliability of modern onshore and offshore wind turbine systems. The data and analysis help to advance understanding of failure rates and development of maintenance strategies. The paper could be improved by providing greater detail to describe the population of turbines in the dataset (without breaching confidentiality requirements). Specifically, it would be valuable to include figures or tables that show the distribution of rated capacities among onshore and offshore turbines, the distribution of turbine ages (operational years), and the shares of technology type such as geared/direct drive/hybrid.

Thanks very much for your comments. We agree that, in an ideal world, we would be a lot more open about the characteristics of the dataset we have used. However, our hands are largely tied by strict confidentiality constraints. We have nevertheless extended Table 3 to partially address the reviewer's comment:

**Table 3: Information about the data sets which have been considered in the analysis.**

	<b>Offshore</b>	<b>Onshore</b>
WT operational years considered	1755	2489
Number of WT OEMs covered	4	9
Rated capacity considered	Up to 9 MW	
Available data period	2006-2024	
Split pitch system components	Electrical: 48%; Hydraulic: 52%	
Split converter system concepts	Fully rated: 48%; Partially rated 52%	
Split drive train concepts	Geared: 78%; Direct drive: 21%	

In the presentation of results, please provide more discussion of the difference between failure rates per MW and per WT. Comparison of Figs. 1 & 2 would seem to indicate that for some components (e.g. MDA, MDY) the failure rate is relatively constant per WT whereas other components (e.g. MDL, MDX) appear to have a constant rate per MW. It would also be helpful to note the average rating of onshore and offshore WTs.

We have extended Section 4.4.1 with the following paragraph to highlight the implications and limitations of linear scaling per MW: "While intervention frequencies similar to these are observed across offshore wind farms with different WT power classes included in the data sets used for this analysis, readers should note that this implicitly assumes a linear scaling per MW for all subsystems, which is a simplification. We have explored this in more detail in (Walgern et al, 2026; Walgern et al., 2025), and for the converter system in (Anderson et al., 2024; Anderson et al., 2025). In that more detailed analysis, the control system, yaw system, lifting gears and meteorological measurement system show no evident trend with regard to rated power. The rotor, pitch, hydraulic, power generation, generator switching and converter systems do. For more accurate insights into scaling for these subsystems, readers are referred to (Walgern et al, 2026; Walgern et al., 2025)."

The comparison of failure rates across OEMs does not add much value because the OEMs must remain anonymous. A comparison across characteristics that could be identified (such as

drivetrain concepts or generator types) would have more value to researchers interested in using this paper to identify areas for future study or application of new maintenance strategies.

We agree that a comparison across design concepts would be valuable. To avoid confounding their effects with other factors (such as the rated power, commissioning date, age of operation etc.), we analyse this in the paper (Walgern et al., 2026) which uses a more advanced methodology. To highlight this, we have added the short sentence to Section 4.2:

*“For further analysis of turbine design choices, see (Walgern et al., 2026).”*

Regarding the comparison of failure rates across OEMs, we agree that this is not helpful to readers for identifying which OEM has the most reliable turbines. However, it does at least confirm that there are significant differences in reliability across different OEMs, which we believe is helpful in the absence (in this paper) of more detailed analysis. We therefore argue against its removal.

The comparison of results for onshore WTs with previous studies would benefit from more explanation of the differences between turbine populations, for example, normalizing by rated power as you have done for offshore WTs.

We thank the reviewer for acknowledging our attempt for the offshore example. While a normalization-based comparison was partially possible in a qualitative way for the study by Faulstich et al., for which sufficient information on the turbine population was available, the other two studies do not provide enough information to provide a similar comparison as for the offshore case.

Please define major component replacement in the context of this paper. Do you use a size or cost threshold to identify MCRs?

Thanks for this comment, we had not noticed this was not defined. We use no threshold to define an MCR but rather count all maintenance actions listed in Table 4. To clarify this in the manuscript, we have added the following to Section 4.4.2 Major component replacements:

*“In contrast to (Carroll et al., 2016), we do not define an MCR via a cost, size or downtime threshold, but directly label it as a MCR in the maintenance data when one of the following components were replaced.”*

Table 2 is unnecessary, the contents are already present in the preceding paragraph.

The table has been deleted in accordance with the reviewer’s wishes.