Reviewer #1 – Authors comments:

1.) **"The abstract provides a concise and complete summary. Solely quantitative results could be presented more precisely in the abstract."**
- We will add a comment on the final comparison between load estimations and observed failure statistics in the abstract and conclusion.

2.) **"Only a few aspects should be rechecked or maybe reformulated. To mention a few (also addressed and highlighted in attached PDF):"**

"Description of the bathtub curve (line 29)"
- We will reformulate the description of the bathtub curve.

"Division of data set (line 267)"
- Since the number of simulated samples is relatively large, the data division is less crucial. It performs similarly well when using (80/20) data split. We will add a comment on this.

"Discussion / comparison model predictions and failure maps (line 324ff)"
- We will further rephrase the comparison for clarification in the revised paper.

3.) **"Technical corrections are nearly not needed as the language is fluent and grammatically correct. Only line 25 should be checked once more as it seems to have words missing or too many (see attachment)."**
- We will correct it in the revised paper.

Please also find our responses to the remaining comments that were addressed in the PDF:

1) **Figure 7: "does it make sense to have a nearly uniformly distributed wind speed? Normally wind distributions look more like a Weibull distribution"**
- The sampling of the wind speed in the training phase (step 1) can be done using any suitable target distribution as it only influences the model training. This distribution is not used in the final power and load estimations. We have chosen a uniform distribution for the training sample in order to have sufficient amount of samples at lower and higher wind speeds, where the wind shear and wind speed standard deviation have a higher variance. However, when the surrogate model is later applied to the SCADA data (step 4), the inputs are weighted according to the site-specific wind field distributions. We will address this in the revised paper.

2) **L. 315 "would be good to describe the comparison of estimation and statistics a little more"**
- We will describe it further in the revised paper.

3) **L.324: "Not sure if this is clear enough. Check again. Also, check if most failures have really occurred at the border of the wind farm. Maybe connect with prevailing wind direction."**
- We will clarify the discussion on comparing the failure observations with the wind farm estimations in the revised paper.

4) **L.331: How many recorded failures for the 5 years? Is it representative enough?**
- Unfortunately, we are not at liberty to share absolute failure numbers, only the normalised numbers. However, we have added further comments (see point 6).

5) **L. 347: "Is this really true? It seems like there are less turbines at the border which have a failure than the ones which have no failures (Fig 6)"**
- We will rephrase this in the revised paper for clarification.

6) **L. 352: "Quantification would be good --> how many failures per turbine expected in 5 years, how many seen? number enough to make comparison?"**
- We will include a reference value for the expected lifetime in the discussion and also point to the fact that more data will be needed to find a statistically significant trend in the averages of the main bearing lifetime (mean time to failure) per turbine or subgroups of turbines.

Reviewer #2 – Authors comments:

1. **"The language is grammatically correct with the only exception of line C1 WESD Interactive comment Printer-friendly version Discussion paper 25 which can be reviewed."**
- We will correct it in the revised paper.

2. **"In line 170 the symbol for row spacing should be reviewed."**
- We reviewed this in the revised paper.

Please find our responses to the remaining comments that were addressed in the PDF:

1. **L. 128: Please indicate: "...the desired quantity" of what "for assessing performance..."**
- We will indicate it in the revised paper.

2. **L. 142: In a print-out version it is not clear that NTN is a reference and looks more like an abbreviation. Maybe the year could be added.**
- We will correct the reference.

3. **Formula 4: What does Ne stand for in this formula?**
- Ne stands for the number of samples used in one training iteration – we will clarify this in the revised paper.

4. **L. 204: If possible give more recent references for O&M costs.**
- We will add a more recent reference on the O&M costs of offshore wind turbines. We will also replace the reference for the main bearing repair cost with a more suitable one from the same author. However, unfortunately we were not able to find a more recent one.

5. **Figure 7: "The wind distribution does not seem plausible with a occurrence frequency of almost 80 for every wind speed. If the graph is meant to represent something different please explain in the text by one sentence."**

- For training the surrogate model, the input variables can be sampled according to any target function, since it does not influence the power and load estimations of the resulting model. Please note that the uniform distribution from which the samples are drawn does not relate to the actual probability distribution of the wind speed. The probability of each sample is taken into account later in the process. We have chosen a uniform distribution in order to have sufficient training samples at lower and higher wind speeds where the wind shear and wind speed standard deviation have a higher variance. We will address this in the revised paper.

6. **Figure 20: Please consider changing one curve into a dashed line. When papers are published in printed form it is often done in black and white. Blue and red are then hard to distinguish in grey sales.**

- We will correct this in the revised paper.