Revision and answers Schepers et al.

We thank both reviewers for a thorough and critical review and for rightfully pointing out several points where our manuscript can be improved.

Although both reviewers use slightly different words we think that they both address some common criticism. In this document we will start with explaining our answers to this common criticism.

Thereafter we will discuss the remaining specific comments from each of the reviewers.

Common comments

- The agreement between calculated and measured wind data is not considered convincing. We understand this criticism, see Appendix A.
- The paper discusses too many unrelated subjects by which the reviewers get confused. Yes we understand this criticism. We want to overcome it with the strategy explained in Appendix B
- Not all conclusions are definite, some of the research is still work in progress. Our answer to that is found in in Appendix C

Specific comments from reviewer 2

We appreciate the fact that reviewer 2 acknowledges the relevance of our work in general and our efforts to identify various real and complex inflow conditions compared to the standards for load calculations.

<u>General comments</u>: Most of them are covered above. Moreover, we will not refer anymore to the sequel report with additional results. These are just more of the same and all references to them led to confusion. You are right that these results were excluded due to space constraints, but we can do without them.

Line 85: do you mean 365 simulations of 24 hours each, and as such arrive to 1 year?

Yes, we have performed 365 simulation of 24 hours (plus a 2h spin-up period for each simulation). This has been added to the text.

Line 91: could you indicate how the azimuthal resolution compares to the standard IEC turbulence box resolution?

An azimuthal interval of 6 degrees is what we often use in our aero-elastic simulations indeed. This has been added.

Line 91: could you also give an indication of how expensive these simulations are in terms of computational time and resources?

The computation time of the LES runs amounts to roughly 2 days on a cluster with 4 NVIDIA Volta GPUs. This has also been added.
Computation time of the load calculations is much faster than real-time for BEM on a simple laptop. The Free vortex wake calculations are a factor 100-1000 slower (dependent on number of wake points and the wake cut-off length etc).

Figure 3 and Figure 4 are not referred to in the text.

– include reference to figure 2 also on line 101 and further around line 110?

– should fig 3 and fig 4 be referred to around line 110-130?

These figures have been removed and replaced by a new figure that shows all the relevant case profiles in one figure with several panels.

Figure 6: to which grid size is the label GRASPref referring to?

This figure has been removed, and we will discuss the validation along the lines given in Appendix A.

Paragraph at line 180: I am not sure I understand this, are you referring to results from a study you did but have not published yet? Why not include those results, who show better agreement, here instead? That also begs the question, what causes some cases to match better than others?

What we were trying to say is that we found a relatively poor agreement between calculations and measurements for the strongest LLJ from figure 6 but the other LLJ cases from the same year are predicted better. However, in Appendix A it is explained that we will describe the validation in a different way and the figure has not been included in the revisions. Instead, we have presented the average low-level jet profiles over the year for the observations, GRASP and ERA5.

Line 185: what is an Excel polynomial curve fitting? I assume you just have used an n-th order polynomial to fit something by minimizing a least-square cost function or something?

Actually, this is a built-in smoothing function (*indicated through scatter with smooth lines*) which is not documented by MS but appears to be some kind of Catmull Rom spline. In the revision the figure is deleted anyhow

Lines 245-250: it is a very interesting teaser to this sequel report, but why not include the results here? As a reader I get the impression the authors already done the necessary analysis.

These sentences might confuse the reader indeed. What we were trying to say is that the low turbulence intensities from GRASP which go together with a LLJ are largely responsible for the low loads where these low turbulence intensities are believed to be true.
We have revisited this entire section. We have included a figure (new version figure 6 bottom panels) that shows the wind speed standard deviation against wind speed average for the observations and GRASP conditioned on the occurrence of a LLJ. This figure shows that LLJ indeed coincide with low TI levels and that observations and GRASP are in agreement about this. We trust this whole point is now presented much clearer.

Figure 12 is confusing to me, it took me a while to figure out that TI, TKE, shear, ... on the titles of the upper row referred to the "TI: case at which extreme TI has been observed". At first this is not obvious because TI, TKE are the same labels used on the x-axis of the respective TI, TKE row plots.

We have updated figure 12 (figure 4 in our new version). We understand it is a lot of information to present, so we have made an effort to clearly explain it in the text and in the figure caption. Nevertheless, we feel it is an important figure because it summarizes our extreme case selection and the comparison of the model with the observations.

figure 13: it is clear that the observed 1Hz equivalent loads (EQL) are lower, but strictly speaking that is only half the story. The other half is the frequency/distribution at which all the different events occur and together they result in the life time equivalent load. You could consider pointing that out as well.

Yes, we fully agree. Using an EQL's conceal the underlying RFC and frequency information. This has been explained in the revision

Regarding: 6. Comparison between aero-elastic loads at extreme events with loads from the reference spectrum:

I assume the extreme loads from DLC1.2 are based on at least 6 seeds per wind speed at 3 different yaw inflow angles (-8, 0 and 8 degrees)? In contrast, only 1x 10 minute realisation is used for the extreme events since they are directly based on a reconstructed/measured inflow field. I understand that in this context the authors want to demonstrate that extreme inflow conditions based on observations/high-fidelity simulations does not necessarily lead to higher loads when compared to the standards. I agree that is an important observation. However, I think it remains plausible that you could in theory create very similar inflow conditions as you have observed leading to extreme loads exceeding the DLC1.2 reference values. Since there is so much uncertainty and variation in turbulence levels across the rotor plane with smaller local "bursts" driving an outlier event (for example: https://dx.doi.org/10.1002/we.497). The more simulations you consider, the higher the likelihood you bump into such an event. From that perspective comparing extremes is only sensible when considering many 10 minute realisations. You could consider stating even stronger in the conclusions that many more aspects require more analysis, and that this study contributes to that bigger picture.

Your assumptions are right: We used 6 seeds for the three different yaw angles. We will point that out in the revision. And yes, more realizations are needed to get a better picture of the extremes. We will also add this to the conclusions.

Comments from reviewer 1

For conciseness, we paraphrase the reviewer comments (*in italic)*:

<u>General comments:</u>

Messy formatting.

Our sincere apologizes, you are very right. Somehow Word is not doing what it is supposed to do but we have taken an effort to solve this as good as possible. Also apologizes for chapter 6 which occurs twice indeed. The chapter number for the conclusions should be 7 Chapter number

Two different topics in one paper.

See Appendix B for an elaborate view on this and how we implemented that in our revision.

Points 1) 2) and 3): Agreement between calculations and measurement is not so good. (our paraphrasing of the comments)

See Appendix A for an elaborate answer to this

4) Acronyms.

Apologies and yes we will spell them out in the revision. EQL is the equivalent load indeed. I remember this abbreviation was often used in the 1980’s when this concept was introduced. Nowadays DEL may be more commonly used indeed.

5) and 6) about 'special cases' and normal production cases

Yes you are fully right, the ultimate loads generally do not come from the DLC1.2. What we are trying to say is that the ultimate loads from the extreme GRASP cases are lower than those from the load set based on DLC1.2 only. But then they should definitely be lower than the ultimate loads from the full IEC load set because the full spectrum considers more cases and the ultimate loads can only become higher by adding load cases.

7) Not a surprising result that feeding the turbulent fields of GRASP leads to lower loads because they have much lower turbulence than the standard class IA conditions.

We respectfully do not fully agree with the conclusion that this is not a surprising result. Perhaps it is not surprising (this is of course up to a reader to decide for itself) but relevant, nonetheless.

We have taken the turbulent wind fields belonging to the most extreme cases from a year of LES results. We have also shown that our LES is able to capture the overall wind statistics at the site well. So, it is reasonable to assume that we produce realistic turbulent wind fields. We computed the loads based on these realistic turbulent wind fields and observe that they are lower than the standard IA class conditions would have given.

A possible explanation is, indeed like the reviewer suggests, that our cases come with TI levels that are lower than the standards. But this is just what the GRASP model (validated by the observations) tells us. The LLJ (and extreme shear) go hand in hand with low turbulence intensity. We fully agree that this is not a new insight but it is less trivial that the complete turbulent structure as modeled by LES, in a particular event like a LLJ with a non-trivial wind speed profile, the loads are reduced. As mentioned above, LLJ's are often considered negative in terms of loads. We have added another picture that shows that both GRASP and the observations have similar TI - LLJ statistics.

8) Elaborate on the dynamic controller

You are right that details on the control are not given. The controller is believed to be representative since it has been designed with standard control design tools available at the partners in the INNWIND.EU and AVATAR projects with which they often design controllers for industry. Below rated wind speed, the turbine controller aims for maximum power production with variable rotor speed operation using a speed dependent generator torque setpoint (for optimum tip speed ratio) and constant optimal blade pitch angle. Above rated wind speed, the rotor speed and generator power are regulated to their nominal rating using constant generator torque and collective blade pitch control. We have included this information in the revision.

9) Possible other discrepancies.
We use the AeroModule which is a single code with an easy switch between the two different aerodynamic models: BEM and FVW. This assures the rest of the input (geometrical, structural, aerodynamic blade data, turbine data, control algorithm) to be the same. So yes we are sure that the differences are caused by the aerodynamic models. They are consistent to conclusions from the AVATAR project and although differences are large indeed the 14% difference is not larger than the differences found in the AVATAR project. In the final report from AVATAR it is written that:
Comparisons between aero-elastic calculations based on BEM showed a 15% over-prediction in fatigue loads compared to those from FVW, probably due to an inaccurate prediction of time-varying induction effects.
See also Appendix C

10) Table was unclear
The deterministic component is the (mainly) the shear driven component. In the revised version of the paper we call it the azimuthally binned averaged component, which in the case of a linear system is similar to the deterministic component.

11) Unclear paragraph on harmonics

Yes I found it difficult to explain the 1P and 2P behavior. We have reformulated it as follows which I hope is better understandable

…. Some further explanation is offered by Fig. 15. This shows a comparison between the azimuthally binned averaged flatwise moments for the LLJ and DLC1.2. Azimuth angle zero indicates the 12 o clock position. The rotor rotates clockwise so azimuth angle 90 indicates the 3 o clock position when looking to the rotor. The variation from DLC 1.2 has a 1P variation with a relatively large amplitude. This is the behavior of the flatwise moment in an atmosphere with 'common' vertical wind shear. The wind speed (and so the loads) decreases when the blade rotates from the vertical upward 12 o'clock (zero azimuth) position to the vertical downward 6 o'clock (180 azimuth). The flatwise moment increases again when the blade rotates from 180 degrees towards 360 degrees.

The variation in flatwise moment from the low level jet is very different. It shows a 2P variation with a relatively small amplitude. This 2P variation can be explained with the LLJ wind speed profile from Fig. 2 which shows the wind speed to be low at 0 degrees azimuth (the 12 o' clock position, when the blade is pointing vertical upward) and at 180 degrees (the 6 o' clock position, when the blade is pointing vertical downward). The wind speed is maximum at (approximately) hub height which correspond to azimuth angles of 90 and 270 degrees (i.e. the 3 o' clock and 9 o' clock position when the blade is standing horizontally)
This velocity variation is reflected in the flatwise moment. It is low at 0 degrees, high at (roughly) 90 degrees and 270 degrees and low again at 180 degrees. This leads to a 2P variation but the load amplitude is relatively small. Hence although the 2P load variation happens twice as often as the 1P load variation from the DLC 1.2. the lower amplitude of the variations lead to a lower fatigue

12)Work in progress and no novel conclusions

We refer to our comments above and in Appendix C.

Appendix A: Agreement between calculations and measurements

We agree that the correspondence of a some of the presented 10-minute average wind speed profiles may be considered poor considering typical yearly averaged error metrics (modeled vs. observed wind speed) of a mesoscale or reanalysis dataset. However, we would like to point out that this can be expected when focusing on extreme cases, which are, by definition, the outliers in a dataset. The reviewer comments have made us realise that, in fact, a one-to-one comparison of extremes is not the most relevant. What is more relevant is the question: does GRASP capture, in a climatological sense, the (extreme) wind characteristics? In our revision of the paper, it is exactly this approach that we have taken: focus on the validation on the distribution (occurrences and magnitudes) of extreme wind events rather than a point-by-point comparison of single extreme events.

However, the reviewer comments were already indicating that the paper is trying to convey two messages in one paper. So, although an elaborate validation of extreme winds is an important and necessary next step that needs to be performed and presented in detail, we feel that putting too much emphasis in this would create a severe imbalance in the contents of the paper. To find the right balance between the validation of GRASP winds and the actual aero-elastic load calculation, we intend to include a number of extra figures showing the distributions of wind and turbulence and

90th percentiles of strong veer and shear events and to remove the one to one comparisons of the profiles. Furthermore, we give more emphasis to further GRASP validation in our recommendations for future work.

Appendix B: Too many messages in one paper.

The paper addresses too many messages which confuses the reader.

Yes, we understand and that is why we had added section 2.

In section 2 it is mentioned that, apart from demonstrating the combined GRASP-PHATAS tool, the research aims to investigate the impact of extreme events on the load spectrum but also the accuracy of a standard BEM model for the calculation of these events. These are two different subjects indeed which are to some extent unrelated. However, in our case the second subject is a spin-off from the first subject in view of the fact that we performed our calculations with AeroModule which is a code based on two different aerodynamic models: A BEM based model and a higher fidelity FVW model. So, insights on the accuracy of BEM are automatically obtained. We then feel it is a pity not to share these insights and this is supported by the comment from the second reviewer who writes that the consideration of BEM with disc averaged induction needs attention.

To make a long story short: We prefer to keep most of the results in the paper but we have prepared the reader in section 2 better on the fact that results are discussed on two unrelated subjects by explicitly mentioning this.

Another step with which we have taken to improve readability is to divide section 6 in two subsections:

- Assessment of loads from extreme events
- Accuracy of calculating loads from extreme events.

This is in line with the recommendation from the first reviewer

In order to improve readability, we have skipped the result for a rigid rotor and the discussion of flexibility effects. In retrospective we believe that these results and the associated discussion are less relevant and they only add to the confusion.

Appendix C Conclusions are preliminary and not new:

Our main take away messages are:

- The extreme events and in particular the LLJ's which we have considered do not give loads which exceed the loads from a standard design load spectrum. We think this is new because it is often stated that LLJ's have a negative effects on loads, see e.g. Duncan, J. November 2018 . But it is a preliminary conclusion indeed. We agree that the methodology needs further validation and we need to consider more aspects (more seeds, more years etc). We tried to make that clear by saying *that more research is needed to warrant a conclusion, especially in the validation of the on-site turbulent wind fields*. So, we agree that this is a preliminary conclusion indeed but we are not sure how to make this more explicit.

- BEM overpredicts fatigue loads for shear driven cases. We are confident on this conclusion. Indications for this overprediction have also been found in other studies, we refer to the EU project AVATAR and also the recently finished Dutch national project Vortex Loads

(Boorsma, K., Wenz, F., Aman, M., Lindenburg, C., & Kloosterman). However both AVATAR and VortexLoads consider artificial load cases where the present study is based on extreme cases which result from a physical LES model. So, we think this is a new conclusion.