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.

General comments: 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: Yes, we have performed 365 simulation of 24 hours (plus a 2h spin-up period for each simulation).

Line 91: An azimuthal interval of 6 degrees is what we often use in our aero-elastic simulations indeed.

Line 91: The computation time of the LES runs amounts to roughly 2 days on a cluster with 4 NVIDIA Volta GPUs. Computation time of the load calculations is much faster than realtime 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).

Figures 3 and 4 are implicitly mentioned on line 106 but it is much better to follow your suggestions and refer to the specific figures indeed. We will do that in the revision

Figure 6: Graspref: This figure will be removed, and we will discuss the validation along the lines given in Appendix A.

Line 180: 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, see figure 7. We could reformulate it follows: *Figure 6 shows a rather poor agreement for this particular LLJ event. However there are other LLJ events which are predicted much better, see figure 7.*

However, in Appendix A it is explained that we will describe the validation in a different way and the figure will not be included in the revisions.

Line 185 Excel polynomial fitting: 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 will be deleted anyhow

Line 245-250. 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 will reformulate it as follows:
In the sequel of the report it will be shown that the loads from the LLJ are relatively low. The low loads at a LLJ are partly caused by the very low turbulence intensities which go together with an LLJ. It is then important to know whether these low turbulence intensities at LLJ's are also found in the measurements (as a matter of fact the measured turbulence intensitities are even lower).

Figure 12: Yes, you are right but we will eliminate this figure.

Figure 13: Yes I fully agree. Using an EQL's conceal the underlying RFC and frequency information. We will explain that in the revision

Regarding 6: 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

General

- Messy formatting. Our sincere apologizes, you are very right. Somehow Word is not doing what it is supposed to do but we will solve it. Also apologizes for chapter 6 which occurs twice indeed. The chapter number for the conclusions should be 7
- Two different topics in one paper. See Appendix B

1) 2) and 3) Agreement between calculations and measurement is not so good. See Appendix A

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) 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) this is indeed what we are trying to say. But it is the LLJ (and extreme shear) which give this low turbulence intensity. The fact that a LLJ and extreme shear give a low turbulence intensity is in itself not so surprising because we know that a high shear goes together with a low turbulence level. But it is important to realize this low turbulence intensity at LLJ's goes together with a low turbulence

level by which the loads are reduced. As mentioned above, LLJ's are often considered negative in terms of loads. We are planning to add another picture that shows that both GRASP and the observations have similar TI - LLJ statistics.

8) 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 will include this information in the revision

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

11) yes I found it difficult to explain the 1P and 2P behavior. I will reformulate 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

11): See 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 intend to take: 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 intend to give more emphasis to further GRASP validation in our recommendations for future work. As an example, we include a few of the plots we intend to include in the revision below. Note that these are preliminary and not the final format or quality.

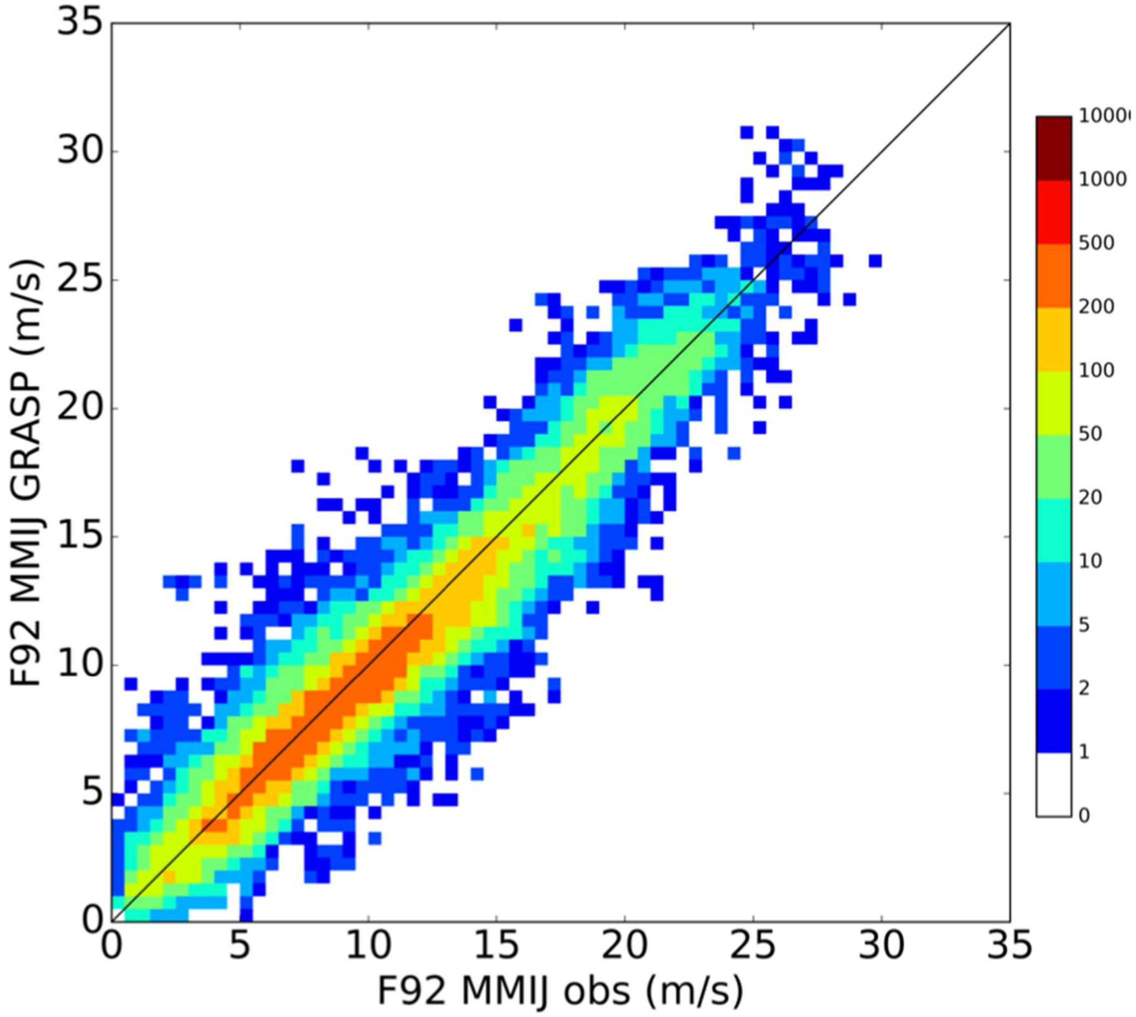
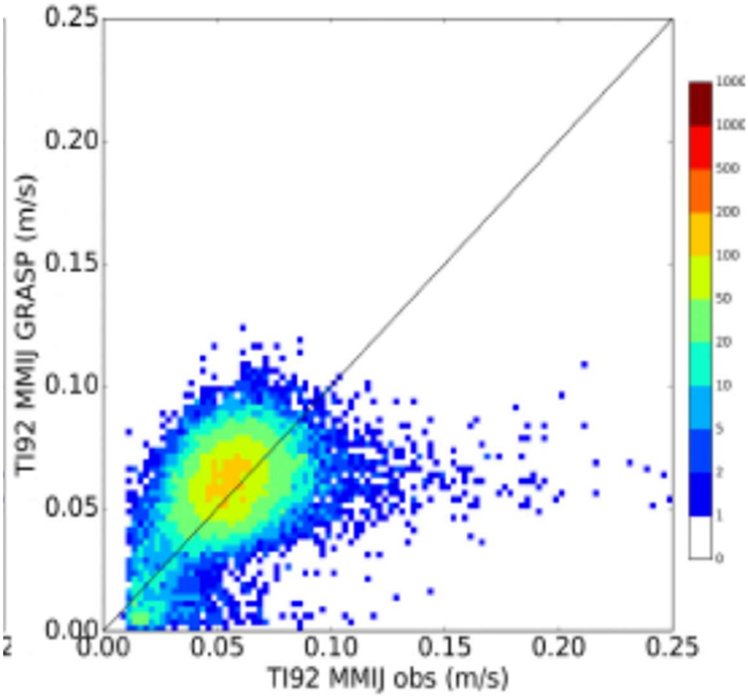
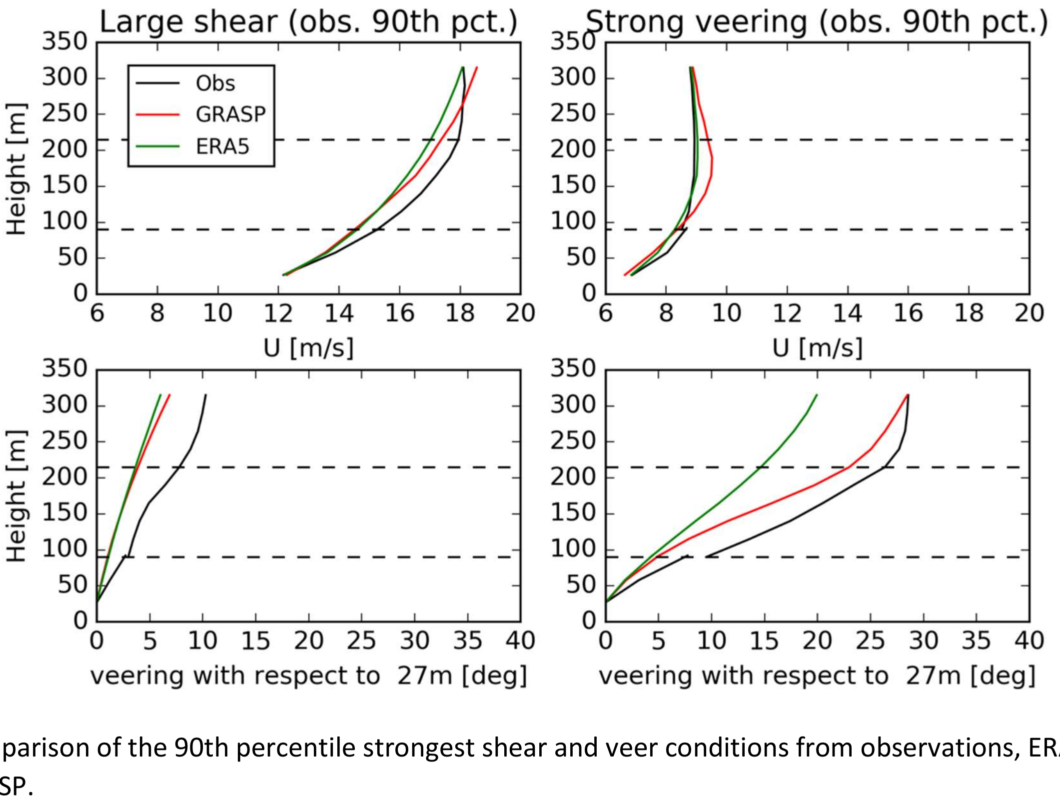


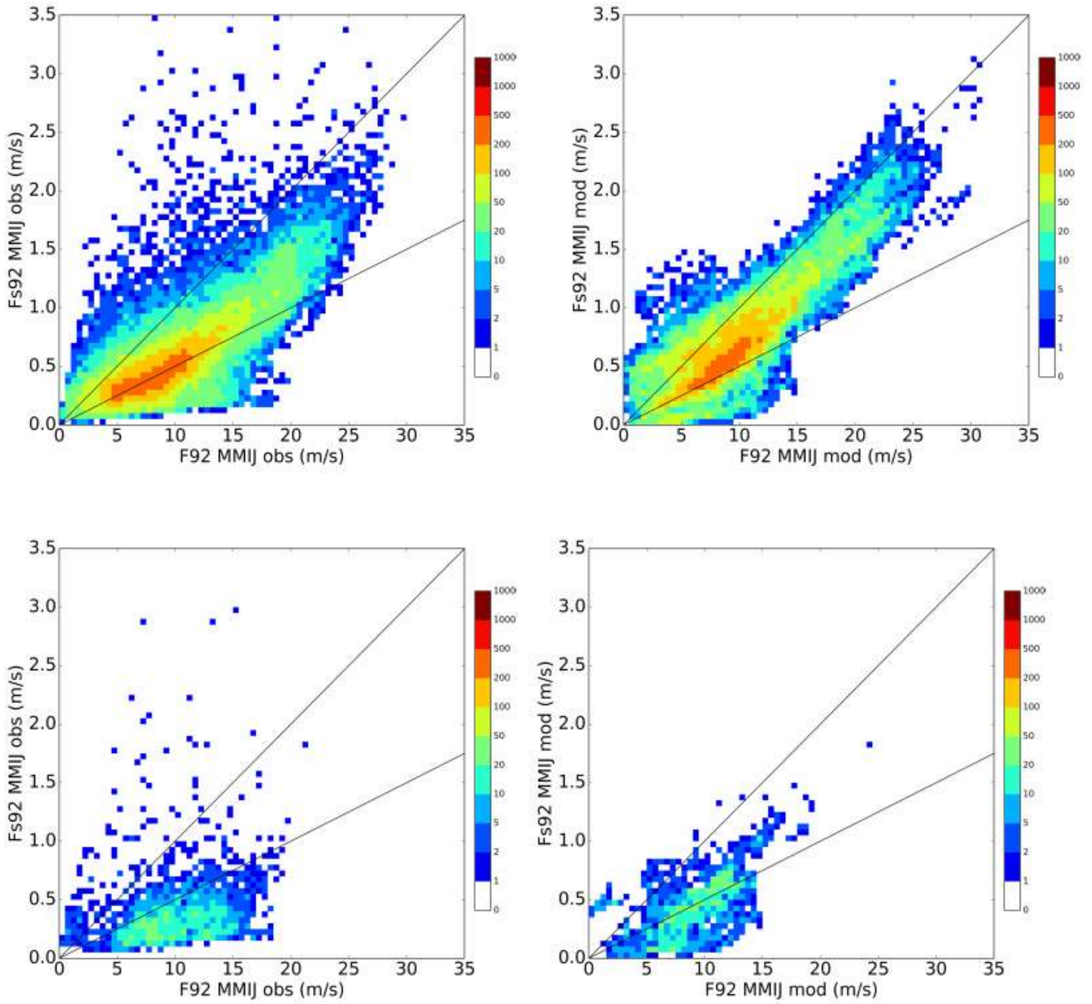
Fig: Scatter density plot of modeled vs. observed 92m wind speed.



Scatter density plot of modeled vs. observed TI at 92m height.



Comparison of the 90th percentile strongest shear and veer conditions from observations, ERA5 and GRASP.



Standard deviation of the wind speed versus wind speed at 92 m for one year of MMIJ data from observations (left panels) and GRASP (right panels). Top panels represent all data, bottom panels only include data records for which either the observed or the modeled wind speed profile is classified as low-level jet (5.6% of the data). For reference, the solid lines indicate TI values of 10% (upper lines) and 5% (lower lines)

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 will prepare 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 will 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 will also skip 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 donot 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.