Review of "Aero-elastic load validation in wake conditions using nacelle-mounted lidar measurements"

Summary: Upstream measurements of nacelle-mounted Doppler lidars are used to characterize the inflow wind field of a wind turbine and to set-up simulations for load and power predictions. These predictions are then validated by
(i) quantifying uncertainty indicators between lidar-based prediction and measurements of on-board sensor in waked conditions
(ii) comparing uncertainty indicators quantified between mast-based prediction and on-board sensors in free wind conditions with the results of (i).
Lastly, the sensitivity of the results to the input parameters of the load simulation is investigated. It is concluded that using lidars for load and power validations is a viable possibility, but further research is needed to compare with IEC standards.

General comments: The manuscript motivates the research question and its relevancy. The methods describe the measurement site, the scan set-up of the lidars, and their processing in sufficient detail, but is scant with information on the simulation and the on-board sensors. My main issues with manuscript are (explained in more detail in the specific comments below):

1. The same inflow parameters that are used to characterize a free-stream inflow are applied to a waked inflow without modification. This includes the assumption of a power-law wind profile that is not valid within a wake.
2. Section 4.3, which presents the validation, is missing structure and components of the validation are not defined. Further, it is not clear to me, why the mast was chosen for the wake-free reference case and not lidar, because this adds another variable in the interpretation.
3. The on-board sensor and the simulations are introduced in only one sentence with a reference to another paper. Since they are as integral as the lidars for the validation, they should receive more attention in my opinion.

I classified this as major revisions, because the first issue could change the results or modify the research questions and conclusions, and the other comments point at an incomplete manuscript.

Language: I noticed only a few typos or grammar mistakes in the manuscript with the disclaimer that I am not a native English speaker. Parts of the manuscript would benefit from structuring with subsections and paragraphs (sections 1, 3.4, 4.1, and 4.3 specifically).

Specific comments

- Page 1, line 8-9: This sentence reads to me, that lidar-based load predictions in waked conditions are compared against the mast-based predictions from a mast located in the free wind at the same time. But I believe that the intended meaning is, that the uncertainty of lidar-based predictions against on-board sensors in waked conditions are compared with the uncertainty of mast-based predictions against sensor data in free wind conditions.

- Page 1, line 10: Why is only the smallest increase of the relative error given? What was the largest observed increase?

- Page 1, lines 10-11: How do they impact the predictions? (e.g. do low wind speed lead to a better uncertainty of the prediction or the opposite?)

- Fig. 2: The remainder of the manuscript often uses x/D as the upwind distance relative to the turbine with the lidars (opposite to here, where it is the downwind distance from the turbine that causes the wake).

- Page 5, eq. (1): Within the wake, the wind speed profile is not following a power law profile. Therefore, I am not convinced that the shear exponent resulting from a partially or fully waked inflow is meaning full. Since the inflow parameters are later used as input for the simulations, the simulated conditions can be expected to be very different from the conditions experienced by the wind turbine.
For the CW lidar it seems possible to retrieve spanwise fields of the longitudinal mean velocity provided with some interpolation. Is it possible to initialize the simulations with them instead? As it is, the approach would be better suited to answer what errors are entailed by applying procedures developed for free stream conditions to waked conditions.

- Page 6, line 2: The simulations should be introduced in more detail.

- Page 7, line 10: I assume that the assumption of homogeneity is referring to horizontal homogeneity only and not including vertical homogeneity?

- Fig. 4: Why is the cup anemometer and not the sonic anemometer used for the turbulence measurements? I would expect that the standard deviation from a sonic anemometer is better since it is not affect by cross-contamination and inertia.

- Eq. (14): Why are only the positions B1 and B2 in the upper half of the rotor considered and not the beams B3 and B4 in the lower half?

- Page 12, line 9: How was it determined that the mast is wake-free?

- Fig. 6: As I understood from section 3.1, the effects of the induction zone were removed from the lidar measurements in the wind field reconstruction and the text mentions that the left panel shows reconstructed mean wind speeds (page 12, line 12). Therefore, I am wondering why the effects of the induction zone are present in the lidar data or whether the shown data is based on eq. (5) and not eq. (6)?

- Fig. 6: I am confused, because the axis labels state the ratio between lidar measurements and mast measurements, but the caption states the slope of a linear regression between.

- Page 14, lines 12-14: From the text I understand that for each wind speed bin an ensemble-averaged spectrum was computed and to each of those spectra the model is fitted to estimate L (i.e. for each wind speed bin a separate L is computed). However, that does not line up with single spectrum presented in Fig. 7 (left) and three length scales reported in line 24.

- Page 14, lines 17-18: Since the shown spectra are normalized with their respective variance, I don't see this from Fig. 7 (left) and only from Fig. 6 (middle).

- Page 14, line 29: At a given wave number, the turbulence kinetic energy depends on the absolute value of energy spectrum and not its slope. Therefore, I am not understanding this sentence. Also, the term rotor sampling frequency could be explained, because I could not find it in quick search and I am not familiar with it.

- Page 15, line 2: How many bins do you have or what is lowest amount of samples in a bin?

- Page 15, line 16 to page 16, line 8: The overall validation approach seems sensible to me. However, in my opinion, it is not clearly written down and has to be pieced together from two

different places in the manuscript. Separate subsections for the power and the two bending moment might be help to make it easier to digest.
In particular, precise definitions of $\tilde{y}$ and $\hat{y}$ were not provided in this manuscript (I looked at the definitions and explanations in Dimitrov et al. (2019) and hope they are also valid here). Also, the generation of two separate bin averaged wind speed ensemble from the previous section should be recalled here. The difference between the mean and the ensemble average with respect to this data set should be explained explicitly. I believe that an extensive rework of section 3.4 is needed, because I misunderstood an essential part of the validation on my first reading and I believe that was not entirely my own fault.

- Page 15, line 18: Check equation. The square should be outside of the bracket.

- Page 15, line 19: Check equation. Missing a square.

- Page 17, line 8: I am not yet understanding why this bias ratio was chosen as an indicator for the behavior of the uncertainty. With a mast-based reference case, there are two influencing factors with (1) the differences between mast vs. lidar and (2) waked vs. free wind conditions. Why not use lidar-based predictions for free wind conditions as the reference? Then only waked vs. free wind conditions remain, which would make interpretation easier.
Also, in later occurrences of the bias ratio it is not stated whether it is an over- or underestimation (I am assuming that both 1.02 and 0.98 would be given as 2%).

- Page 17, line 20: The section is already quite loaded. It might be worth to consider to separate the filtered and unfiltered comparison from the rest of the section. The same might be considered for the length scale.

- Page 18, lines 11: I believe it should be "the increased effect".

- Page 18, lines 12: I believe it should be "higher".

- Page 18, line 15: Why is the effect of L in the sensitivity analysis minor, but it had an effect of up to 15% in the previous section (Table 2)?

- Page 22, lines 10-11: Consider rephrasing this sentence, because it can be understood in two ways.

- Page 22, lines 27-28: Reiterating a previous comment, I am not convinced that the power law should be applied within the wake, because the wind profile has a different shape. The conclusion regarding the vertical wind profile here is based on the sensitivity to the shear exponent and I am not convinced that it holds for the same reason, because the simulated wind fields might have been very different from the real wind field.
Also, I am confused regarding the horizontal wind profile, because I cannot remember that horizontal gradients were accounted for in the wind field retrieval. Is that referring to the wind veer maybe?

- Page 29, lines 23-24: The reference of Dimitrov et al. (2019) seems to be out of place and should appear after Dimitrov et al. (2018) assuming the sorting hierarchy is first author alphabetical followed by year.