Author response to reviewer 2

The authors response is shown in red
Changes implemented in the new version of the manuscript are shown in blue

We thank the reviewer for the comments, which we consider very important. Here our response to each of them.

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 onboard 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.

We agree that the power law is not valid in wake conditions. The vertical wind speed gradient in wakes can be seen as the combination of the atmospheric shear, which can typically be described by a power-law profile, and the wake deficit. However, due to the complexities of wakes, i.e., horizontal and vertical meander and, at NKE in particular, scenarios with multiple wakes the velocity gradients are very complex. It would be very complicated to separate each effect from lidar measurements, without tracking the wakes' center locations in time and without knowing the wake-free atmospheric shear. With the current setup, the derived wind characteristics anyway minimize the error between the modelled wind field and lidar measurements. Fitting a power law on these data is equivalent to applying a low-order approximation to a complex nonlinear function, effectively there is minimal difference between using the average value and the power law. As shown from the the sensitivity analysis in Fig. 10, the shear exponent does not have a significant impact on the predictions. The results from Table 2 and Table 3 as well as those in Fig. 10 and Fig. 11 indicate a higher load predictions uncertainty in wakes compared to free wind conditions. As we assume that the observed deviations in load predictions are solely due to the error in the wind field representation (see Page 3 line 20), it can be inferred that the increased uncertainty is due to the flow modelling assumptions of the used wind field model. This was stated and discussed in the discussion section (see Page 21 line 9-19). In addition, we have made changes in multiple relevant parts of the manuscript, to outline the caveats of using a power law together with wake deficits.

We think that an improved scope of the work is required as well as an extended discussion regarding the limitations due to the modelling assumptions. We therefore replaced or added text:

Page 1 line 4:

From: "We account for wake-induced effects by means of wind field parameters commonly used as inputs for load simulations, which are reconstructed using lidar measurements."

To: "The wake flow fields are described by lidar-estimated wind field characteristics, which are commonly used as inputs for load simulations, without employing wake deficit models."

Page 1 line 13

From: "Finally, the experiences from this study indicate that characterizing turbulence inside the wake as well as defining a rotor equivalent wind speed model are the most challenging aspects of load validation in wake conditions."

To: "Finally, the experiences from this study indicate that characterizing turbulence inside the wake as well as defining a wind deficit model are the most challenging aspects of lidar-based load validation in wake conditions."

The following text was added in the introduction on page 2 line 21:

Based on these findings, we extend the load validation procedure defined in Dimitrov et al. (2019) to include wake conditions. Therefore, wake-induced effects are accounted for by means of wind field parameters commonly used as inputs for load simulations, which are reconstructed using lidar measurements, yet without employing wake deficit models. The objective of this study is to demonstrate how loads in wake conditions can be predicted accurately, quantify the uncertainty, and compare it to the uncertainty of mast-based load assessments in free wind.

The following text was replaced in the the discussion on page 21 lines (9-19)

From: "Although the present work does not focus in details on the performance of the two lidar systems, the findings indicate that the main sources of uncertainty in load predictions are related to flow modelling assumptions. Power production levels are highly dependent on the estimated mean wind speed at hub height. The observed power prediction's deviations, in both partial- and full-wake situations, indicate an inaccurate reconstruction of the mean wind speed. More precisely, the flow modelling assumptions, including horizontal homogeneous wind flow, power law vertical wind profile and linear veer within the scanned areas, introduce larger errors in wake than wake-free conditions. Furthermore, we do not distinguish situations where the lidar beams are partly measuring inside the wake and partially outside to reconstruct the inflow wind field. This could be resolved if wake characteristics as shape, depth and center position are integrated in the WFR techniques (Trujillo et al., 2011). Deriving a rotor equivalent wind speed model, which accounts for velocity gradients as well as wake characteristics is necessary to improve the accuracy of power and load predictions."

To: "Although the present work does not focus in details on the performance of the two lidar systems, the findings indicate that the main sources of uncertainty in load predictions are related to flow modelling assumptions. The wind velocity gradient in the wake is characterized by the combined effect of the atmospheric shear and the wake deficit. The former can be explained by a power law profile, while the latter is often approximated in the far wake by a Gaussian function, whose depth and width depend on ambient conditions and turbine operation regimes. Further, the 10-min average wind velocity gradient, observed from a fixed point, will be largely influenced by wake meandering in the lateral and vertical directions increasing the complexity of the velocity field in the wake region. The results of the power prediction's deviations in Table 2, in both partial- and full-wake situations, indicate a less accurate reconstruction of the wind field when compared to wake-free conditions. Although we demonstrate a low sensitivity of the loads to the shear exponent for all the analysed sensors (see Fig. 10), it is envisioned to more appropriately account for wake-affected velocity gradient profiles and determine whether or not this will significantly improve the accuracy of power and load predictions"

The following text is added in the conclusion on page 23 line 9:

Future research should apply a wind deficit model that accounts for the combined effect of atmospheric shear and wake deficit, and quantify the uncertainty of resulting power and load predictions.

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.

We conducted an extensive re-structuring of Section 4.3; more details are given in the specific comments below. The mast is here the reference because it would be unfair to choose as reference any of the two lidars. Further, the current IEC load validation approaches under free-wind conditions are based on measurements obtained from a met mast; lidars are still not recommended for load validation purposes. However, following the suggestion of the reviewer, we now use the lidars as the reference case. The load statistics in wake conditions obtained with one of the lidars are compared to the respective statistics of the same lidar in free wind conditions. We keep the results based on the mast for completeness and show the uncertainty of lidar-based load validation in free-wind conditions.

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.

We added additional details on the on-board sensors in "Sect. 2.2. Measuring campaign" by including the location of the strain gauges on the wind turbine structures (i.e., blades and tower) and the sampling frequency of the load data.

The following text was added on page 4 line 1:

The wind turbine T04 was instrumented with sensors for load measurements at the roots of two blades, tower top, and tower bottom (Vignaroli and Kock, 2016). The strain gauges were installed at 1.5 m from the blade root flange, at 11.85 m below the lower surface of the tower top flange, and at 5.9 m above the upper surface of tower bottom flange. The data acquisition software was set to sample at 35 Hz on all channels. Additional data were provided by the supervisory control and data acquisition (SCADA) system including nacelle wind speed and orientation, power output, blade pitch angles, and generator speed.

We added additional details on the simulations in "Sect. 3 Methodology", including the description of the HAWC2 model and add relative references.

The following text was added on page 6 line 1:

Load simulations are carried out using the state-of-the-art aero-elastic HAWC2 software (Larsen and Hansen, 2007). The structural part of the code is based on a multi-body formulation assembled with linear anisotropic Timoshenko beam elements (Kim et al., 2013). The wind turbine structures (i.e., blades, shaft, tower) are represented by a number of bodies, which are defined as an assembly of Timoshenko beam elements (Larsen et al., 2013).The aerodynamic part of the code is based on the blade element momentum (BEM) theory, extended to handle dynamic inflow and dynamic stall (Hansen et al., 2004), among others. In the present study, the HAWC2 turbine model is based on the structural and aerodynamic data of the Siemens SWT2.3-93 turbine and is equipped with the original equipment manufacturer controller.

References:

Hansen, M., Gaunaa, M., and Aagaard Madsen, H.: A Beddoes-Leishman type dynamic stall model in state-space and indicial formulations, 2004

Kim, T., Hansen, A. M., and Branner, K.: Development of an anisotropic beam finite element for composite wind turbine blades in multibodysystem, Renewable Energy, 59, 172–183, https://doi.org/10.1016/j.renene.2013.03.033, 2013.

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.

The following text was replaced in the abstract on page 1 lines (8-9)

From: "The uncertainty and bias of aero-elastic load predictions are quantified against wind turbine on-board sensor data. We consider mast-based load assessments in free wind as a reference case and assess the uncertainty in lidar-based power and load predictions when the turbine is operating in partial- and full-wake"

To: "We assess the uncertainty of lidar-based load predictions against wind turbine on-board sensors in wake conditions and compare it with the uncertainty of lidar-based load predictions against sensor data in free wind."

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

In order to write a concise and short abstract, we reported the best reachable accuracy of load predictions. However, to show the sensitivity of the error due to inflow wind conditions, we now include the range of the error based on the unfiltered turbulence measures with turbulence length scales as for free wind and waked situations.

The following text is replaced in the the abstract on page 1 lines (10)

From: "Compared to the reference case, the simulations in wake conditions lead to an increase of the relative error as low as 4%"

To: "Compared to the free-wind case, the simulations in wake conditions lead to increased relative errors (4–11%)."

- 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?)

The analysis in Fig. 11 shows how uncertainty varies as function of the inflow mean wind speed. The results show that low wind speeds lead to higher bias and uncertainty of the predictions. We also mention it in the text (see Page 20, line 3). This trend is noticeable for power predictions, as the power curve is steeper at low wind speed compared to the curve above rated wind speed (Fig. A.1 in the appendix shows the normalized power curve of the turbine). Further, as the power levels are function of the cube of the wind speed, a small error in the estimated wind speed can lead to a significant bias in the power predictions, particularly at low wind speeds.

The results in Table 2 and 3 show the impact of both turbulence intensity and turbulence length scale on the load predictions. We discuss this impact in "Sect. 4.3 Load simulation results" on page 17 lines 20–32.

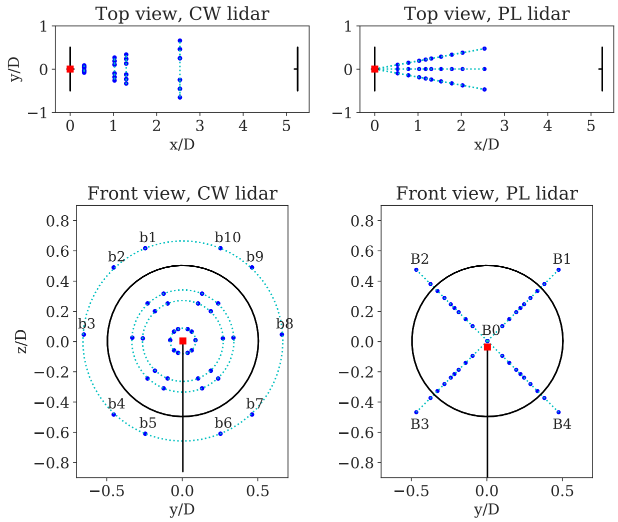
We believe that information regarding how the wind field characteristics impact the load predictions and their uncertainty should be found in the related sections in the manuscript. Following the reviewer comment regarding Sect. 4.3, we did an extensive re-structuring of the section, where we clarify the dependency of the uncertainty of predictions on the inflow conditions (see below for more details).

We address the reviewer comment in Sect. 4.4, and we divide the section as into Sect. 4.4 Sensitivity analysis and Sect. 4.4.1 Uncertainty distribution as function of wind speed,

- 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).

This was corrected. The figure is now replaced by the one below.

- 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.

Please, see response to major comment nr. 1.

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

Please, see response to major comment nr. 3.

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

The sentence relates to Eq. 5. The inhomogeneities, as wind shear, are introduced after the statement in Eq. 6.

- 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.

The cup is mounted at 80 m, which is the hub height. The sonic is mounted at 76 m. Previous work on the characterization of wind conditions at the NKE site that included wind speed and turbulence showed a discrepancy between the sonic- and cup-based mean wind speed of 2.6% and about 12.3% regarding the longitudinal velocity variance (Peña et al., 2017). To reduce the uncertainty of the mast-based and lidar-based wind characteristics, we choose the cup anemometer for this analysis. The same approach was used in Dimitrov et al. (2019).

The following text is added on page 10, line 8.

Previous work on the characterization of wind conditions at the NKE site that included wind speed and turbulence showed a discrepancy between the 76 m sonic- and 80 m cup-based mean wind speed of 2.6% and about 12.3% regarding the longitudinal velocity variance (Peña et al., 2017). To reduce the uncertainty of the mast-based and lidar-based wind characteristics, we choose the cup anemometer at 80 m, which is the hub height, for this analysis.

- 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?

Preliminary work (Peña et al., 2017) showed that the lidar availability highly reduces when using the bottom beams. This was simply because the lenses of those bottom beams (B3 and B4) became dirty due to contamination from the cleaning system of the CW lidar.

The following text is added on page 11, line 3.

Preliminary work (Peña et al., 2017) showed that the lidar availability highly reduces when using the bottom beams. Therefore, we use the top beams of the PL lidar for this particular analysis.

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

Some details were described on page 11 line 19. However, we now add an Appendix to describe the application of the wake detection algorithm to the mast measurements.

Appendix A: Wake detection from mast measurements

The wake detection algorithm (see Sect.3.4) is extended to the mast measurements to classify 10-min periods where the mast is in free or wake situations. For this purpose, turbulence observations from the cup anemometer at 80 m and vertical wind shear computed using the measurements from the cup anemometers at 57.5 and 80 m are used as wake detection parameters. Their $99th$ percentiles are used as conservative thresholds to characterize the limits of the normal range of the site-specific free wind conditions. The resulting thresholds are $TI_{mast,99} = 0.20$ and $\alpha_{mast,99} = -0.02$. If one of the two limits is exceeded within a 10-min period, the mast is considered in wake conditions and shown with green markers in Fig. 1.

- 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)?

We corrected and rephrased the text. In the analysis shown in Fig. 6, we investigate how the lidar-based wind characteristics varies as function of the upstream ranges in free, partial- and full-wake situations. For this particular analysis, we disregard induction effects.

The following text is now added on page 12, line 5:

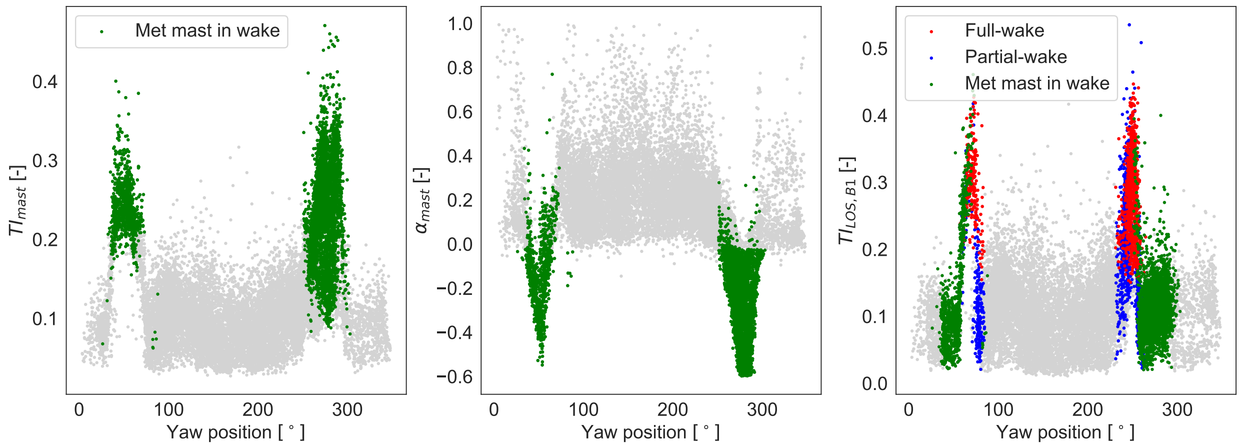


Figure 1: Left and middle: 10-min observations of the turbulence intensity and vertical wind shear at the mast as function of turbine yaw position. Free wind conditions relative to the mast are identified with grey markers, and waked situations with green markers. Right: PL-estimated 10-min wake detection parameter $TI_{LOS,B1}$. Detected wake situations of turbine T04 are shown with coloured markers: wake-free (grey), partial-wake (blue) and full-wake (red). The 10-min periods, where the mast is affected by wakes are shown in green markers.

From: "We observe these effects through the PL and CW lidar-estimated wind speed, turbulence and shear exponent in Fig. 6. Here, the slope of a linear regression model between the free wind mast-measured and lidar-estimated wind parameters in free-, partial- and full-wake situations are shown as function of the upstream distance from the rotor."

To: "Here, the slope (m) of a linear regression model between the free wind mast-measured and lidar-estimated wind parameters in free-, partial- and full-wake situations is shown. In this particular analysis, the lidar-based wind parameters are derived from Eq. (6) evaluated a different upstream distances from the rotor without including induction effects."

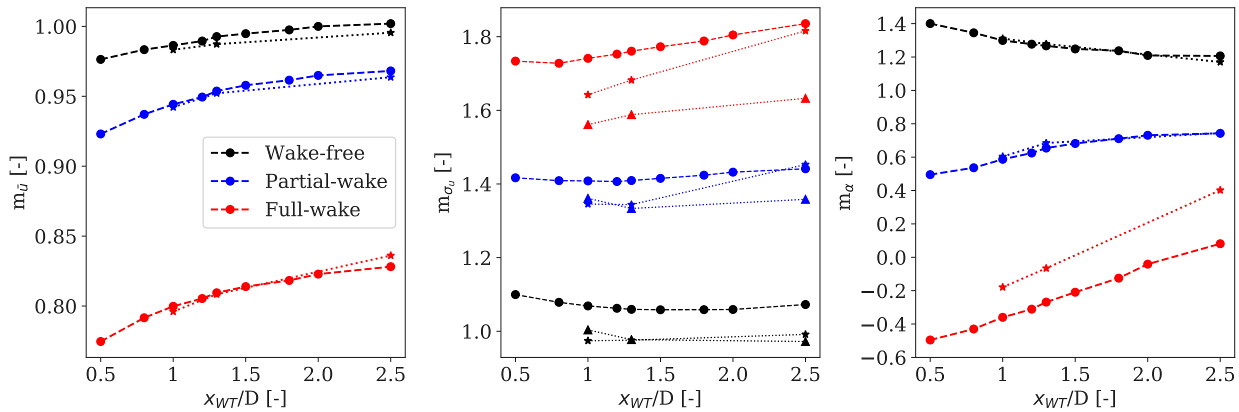
- Fig. 6: I am confused, because the axis labels state the ratio between lidar measurements and mast measurements, but the caption states the slope of a linear regression between.

We changed the figure and provided the corrected axis label.

- 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.

This was not correctly described in the manuscript and we replaced the text accordingly. Also, to improve the readability of the manuscript, we divided section 4.1 in: Section 4.1 Wake effects on reconstructed wind parameters and Section 4.2 Wake effects on turbulence spectra properties.

The following text is now replaced in Page 14, lines 12–14:



From: "The 10-min time series of radial velocity are classified into wind speed bins and the spectra are ensemble-averaged over each wind speed bin. Then, the parameter L is fitted to the ensemble-averaged spectrum weighted on number of samples in each bin.

To: "The 10-min time series of radial velocity are classified into free-, partial- and full-wake situations and the spectra are ensemble-averaged over all conditions within each class. Then, the parameter L is fitted to the ensemble-averaged spectrum."

- 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).

The measured and theoretical spectra are normalized over their respective variance under each condition (wake-free, partial wake and full wake). Without normalization, the curves will be all over the place as the energy content under each condition is very different.

- 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.

We have deleted this sentence as it is confusing. The idea behind this sentence was to explain how the wind field spectra influences the loading conditions on the wind turbine.

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

This sentence in the manuscript may mislead the reader, as we do not classify inflow parameters nor simulation results in wind speed bins. Therefore, we have deleted this sentence.

The following text is replaced on Page 15, lines 1-2:

From: "We ensure close to 500 10-min samples, distributed nearly equally among wind speed bins in the range 4–14 m/s, for free-, partial- and full-wake scenarios."

To: "We select around 500 10-min samples for each of the free-, partial- and full-wake scenarios, which are distributed within the wind speed range 4-14 m/s."

- 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.

We now rename section 4.3 from "Load simulation results" to "Load validation procedure" and re-arrange Section 4.3 as:
- Section 4.4 Load validation procedure
- Section 4.4.1 Power predictions
- Section 4.4.2 Extreme load predictions
- Section 4.4.3 Fatigue load predictions

We also rephrase section 4.4 to provide a better description of the validation approach as:

4.4 Load validation procedure
The load validation analysis is conducted on the dataset described in Sect. 4.3. We analyze about 500 10-min samples distributed between 4-–14 m/s, for free-, partial- and full-wake scenarios. The quality of load predictions is evaluated through one-to-one comparisons against load measurements. The resulting statistics from HAWC2 simulations are denoted by ($\tilde{y}$) and the corresponding measured statistics from the turbine on-board sensors by ($\hat{y}$). Three uncertainty-related indicators are assessed, where the symbol $E(.)$ denotes the mean value and $\langle . \rangle$ the ensemble average.

- Coefficient of determination $R^2 = \langle (\tilde{y} - E(\tilde{y}))^2 \rangle / \langle (\hat{y} - E(\hat{y}))^2 \rangle$
- Uncertainty $X_R = \sqrt{\langle (\tilde{y}/\hat{y} - E(\tilde{y})/E(\hat{y}))^2 \rangle}$
- Bias $\Delta_R = E(\tilde{y})/E(\hat{y})$

The R^2, X_R, Δ_R indicators are computed for free-, partial- and full-wake situations. The 10-min wind turbine statistics investigated hereafter include the mean power production ($\mathrm{Power}_{\mathrm{mean}}$), the extreme loads and 1-Hz damage equivalent fatigue loads of fore-aft tower bottom bending moment ($\mathrm{M}_{\mathrm{xTB}_{\mathrm{max}}}, \mathrm{M}_{\mathrm{xTB}_{\mathrm{DEL}}}$) and flapwise bending moment at the blade root ($\mathrm{M}_{\mathrm{xBC}_{\mathrm{min}}}, \mathrm{M}_{\mathrm{xBC}_{\mathrm{DEL}}}$). Therefore, time-series of 600 s are simulated in the aero-elastic code HAWC2 and load statistics are derived at the location where the strain gauges are installed. A turbulence seed with statistical properties matching those of the measured 10-min conditions is input to the load simulations. The rainflow counting algorithm is used to compute the 1-Hz damage equivalent fatigue loads with Whöler exponent of $m = 12$ for blades and $m = 4$ for the tower. The same approach is used to post-process measured loads. We run simulations using wind field characteristics listed in Table 1, which are derived from both the PL and CW lidars as well as the mast measurements. A more detailed analysis is conducted for partial- and full-wake situations. Here, we investigate how power and load predictions are influenced by filtered and unfiltered turbulence estimates derived in Sect. 4.3, characteristic turbulence length scales derived in Sect. 4.4 and wind parameters derived from lidar measurements at several ranges. We provide detailed scatter plots of measured and predicted load sensors used in the analysis in Figs. B1-B5 in the Appendix. The predictions uncertainties for power production and extreme loads are presented in Table 2, and for fatigue loads in Table 3. We define the lidar-based power and load predictions in free wind as the reference case. Thus, we compare the relative error between the uncertainty indicators derived from wake situations with those from the free wind case. Generally, we observe lower prediction accuracy in partial- and full-wake situations compared to the free wind scenario, while in some cases similar uncertainty levels are obtained. The following sections describe the results in details.

4.4.1 Power predictions

Power production levels are overestimated in partial-wake, but underestimated in full-wake by approximately 4% compared to the free wind case. Larger X_R values are found in full-wake compared to the reference case, although R^2 is above 96%, which indicates a good correlation. We do not observe a significant influence of turbulence intensity levels on power predictions, i.e. by comparing the uncertainties in full-wake between simulations performed with filtered and unfiltered turbulence estimates from the CW lidar in Table 2. In a similar way, small turbulence length scales derived in wakes have a negligible effect on power production levels. The power predictions deviations in partial-wake drop to approximately 1%, when the PL lidar-estimated wind characteristics using measurements up to 1.3 D are used in the simulations. This result indicates the sensitivity of the reconstructed wind field characteristics to the upstream ranges in a strongly inhomogeneous wind field as a partial-wake situation.

4.4.2 Extreme load predictions

The extreme loads ($\mathrm{M}_{\mathrm{xTB}_{\mathrm{max}}}, \mathrm{M}_{\mathrm{xBC}_{\mathrm{min}}}$) are both affected by the turbulence levels as well as the turbulence length scale. We obtain similar deviations in partial- and full-wake as per the free wind conditions, when using unfiltered turbulence estimates and length scales extracted in free wind conditions (see Table 2). However, simulations based on filtered turbulence consistently overestimate extreme load levels (3–7%). The effect of a low value for the length scale is noticeable in full-wake situations, where $L = 7$ m leads to biases of the order of -7% compared to the reference case. Overall, higher X_R values are derived in wakes compared to the reference, while R^2 remains above 89% in all analyzed cases. It should also be noticed that the maximum loads do not increase significantly in wake situations, since the wind speed in the wakes is lower than the free wind (Larsen et al., 2013).

4.4.3 Fatigue load predictions

The biases of fatigue load predictions in partial-wake, using unfiltered turbulence statistics and $L = 35$ m, are comparable with the deviations observed in free wind conditions, as seen in Table 3. The error increases when filtered turbulence from the PL lidar are simulated, leading to an underestimation of fatigue loads between 2–5%. The most significant deviations are observed for $\mathrm{M}_{\mathrm{xTB}_{\mathrm{DEL}}}$ and $\mathrm{M}_{\mathrm{xBC}_{\mathrm{DEL}}}$ in full-wake conditions. The simulations based on filtered turbulence measures and $L = 35$ m lead to an overestimation of blade-root and tower-bottom predictions by 21% compared to the free wind case. The filtered turbulence statistics are predicted with the use of the spectral velocity tensor model and are found to be approximately 11% higher compared to unfiltered turbulence derived from the Doppler radial velocity spectrum (see Fig. 9-middle). The bias

of fatigue load predictions drop to approximately 11%, when unfiltered turbulence measures from the CW lidar are simulated. Overall, extreme and fatigue load predictions show low uncertainty when unfiltered turbulence estimates are used as input in simulations.
Fatigue loads are found to correlate significantly better when a synthetic turbulent field characterized by small length scales is used (i.e. $L = 7$ m). This is demonstrated by improved X_R and R^2 indicators compared to those resulting from simulations with $L = 35$ m. Besides, reducing L from 35 m (free-wind conditions) to 7 m (fitted in full-wake conditions) reduces fatigue blade-root load levels by 15%. The simulations with low length scales and unfiltered turbulence measures provide the lowest deviations in full-wake compared to the reference case, as the error drops to -4% for $\mathrm{M_{xBC_{DEL}}}$, indicating underprediction (see Table 3). These results demonstrate the improved accuracy of load predictions when unfiltered turbulence measures are simulated, and validate the importance of characterizing turbulence spectral parameters for load analysis; as previously demonstrated in Thomsen and Sørensen (1998), Sathe et al. (2012), and Dimitrov et al. (2017).

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

This is now corrected.

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

This is now corrected.

- 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.

Following the suggestion of the reviewer, we changed the reference case, so it is now the lidar-based load predictions in free wind conditions.

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%).

We added the signs to differentiate over- and under-estimation and also specified whether is under- or over-prediction.

- 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.

By re-structuring Section 4.3 into four subsections, as described above, Sect. 4.4.3 Fatigue load predictions will be easier to read.

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

This is now corrected.

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

The first entry “high” describes the fact that we measure high turbulence levels in the wake; the second entry “higher” work as a comparative, in this case between the turbulence levels under wake- and free conditions.

- 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)?

The results in Fig. 10 should be interpreted as the coefficients of a linear regression model between the inputs and the output sensors. The regression models are fitted to a dataset of 850 10-min periods, which are characterized by different inflow conditions (i.e. wind speed, turbulence, shear, etc.) as measured by the PL lidar and the L parameter, which was varied between the defined boundaries (see figure below for the partial-wake). Note that we show the absolute input values in the figure, although the inputs are normalized such that their values are scaled between zero and one, when fitting the regression model. The resulting coefficients give an estimation of the slope of a linear model between inputs and outputs. If a strong dependency is seen between output and inputs, as for the mean wind speed and turbulence in the figure below, a strong sensitivity of output sensor to the input is indicated. In the case of L, we do not see a linear relationship, but rather a scatter plot. This is because, while we vary L, the wind speed and turbulence levels are also varied, and the impact of the latter two is higher to that of L. However, when all inputs are fixed and only L varies, as the results in Table 2 and Table 3, we can estimate the effective influence of L on the predictions.

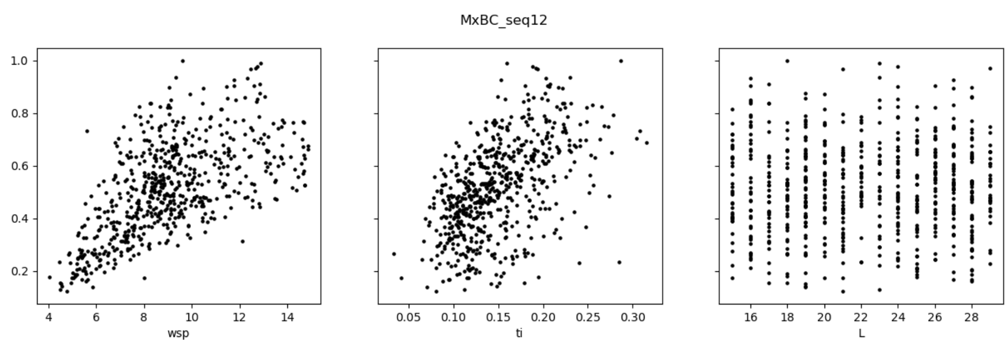


Figure 2: Scatter plot of the normalized fatigue loads at the blade-root as function of inflow wind conditions (mean wsp, turbulence and turbuelence length scale L)

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

The sentence “The effective turbulence approach under predicted fatigue load levels at spacings larger than 5 D (Schmidt et al., 2011).” is now deleted, in case this is what the reviewer was pointing at

- 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?

Please, see the response to major comment nr. 1. We deleted the sentence regarding the horizontal gradient and now discuss the influence of turbulence intensity and turbulence length scale on the power productions levels, as shown below.

The following text is replaced on Page 22, lines 27-28:

From: "Power production levels under wake conditions were strongly related to the mean wind speed at hub height, whereas the vertical and horizontal wind profiles had negligible effects on those levels"

To: "Power production levels under wake conditions were strongly driven by the reconstructed wind speed at hub height, whereas turbulence intensity as well as turbulence length scales had negligible effects on those levels."

- 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.

This is now corrected.