Response to the reviewers for manuscript wes-2021-58

B. Sanderse, V.V. Dighe, K. Boorsma, G. Schepers

September 20, 2021

The authors appreciate the valuable comments from the reviewers in order to improve the quality of the work. In response to that, the author's (on behalf of all co-author) comments can be found below. Note that in the revised manuscript, changes related to reviewer 1 are indicated in orange, while changes related to reviewer 2 are indicated in magenta.

- Reviewer 1
  - RC1: Under steady conditions, Bayesian updating the static airfoil polars is correct as described in the paper. The moment unsteady dynamics (large blade deflections) and turbulent inflow are accounted for, one needs then to also think about updating BOTH the Dynamic stall model AND input static airfoil polars. I suggest that the authors spend a bit of time discussion this issue in the paper.
  - AC1: Thank you for rightly pointing the implications of using the calibrated static airfoil polars for a given case (operating conditions) that involves unsteady flow dynamics. To clarify: we showcase two 'independent' examples - (a) Time independent DanAero MW experiment for calibrating airfoil polars and (b) Time dependent New MEXICO experiment for calibrating yaw model. Airfoil static polars are calibrated using the (time independent) sectional normal force; the cross-validation we show in Figure 9. Since no dynamic (time dependent) data is available to us to perform further cross-validation, the accuracy of the calibrated polars in dynamic flow conditions remain uncertain. If such data would be available, one should realize that the simultaneous calibration of both dynamic stall model parameters and airfoil polars would constitute a very high dimensional problem that might be computationally too expensive. Our approach, in which such effects are separated by using a time-dependent and a time-independent case, is effectively a way to reduce the high dimensionality of this calibration problem.
    We have included this explanation in the Conclusions section. We have also added a sentence in the introduction to stress that the main focus of the paper is on the development of the calibration framework and that the test cases correspond to idealized conditions.
  - RC2: Once the static airfoil polars are updated conditional on the measurements, there is no garantee that the new/update polars are actually correct. One possibility would be to verify the posterior predicted power outpout (if available) or the blade bending moment (if available) with the new/updated polars. I suggest that the authors spend a bit of time discussion this issue in the paper, and if possible compare the power output (for instance) before and after the updating of the polars.
  - AC2: Bayesian calibration of the static airfoil polars is performed using the sectional normal force (measurements). Using the calibrated/updated polars the posterior predictive is plotted along with the measurements and the uncalibrated model results

in Figure 9. It can be observed that the calibrated model predictions clearly overlaps with the mean of the measurements. To obtain the posterior predictive of the power output, or the blade bending moment, given the calibrated parameter values one would need to retrain the surrogate model with the power or blade bending moment as quantity of interest, and use this surrogate model to evaluate the posterior predictive. Alternatively, one could use the full Aero-Module with the calibrated parameters and use that to determine the posterior predictive for the power, but that would be computationally very expensive.
We have added these insights into the manuscript, section 5.1.2.

- Reviewer 2:
  - RC1: It would seem relevant to add the evaluation from the polynomial model to the plots on top of the results from the calibrated and uncalibrated Aero-Module.
  - AC1: Thank you for this comment, we realize that indeed this could cause some confusion. To clarify: we are plotting the evaluation of the surrogate model at both the uncalibrated and calibrated values of the parameter vector. Since the surrogate model is highly accurate (see for example appendix B1), the evaluation of the AeroModule at these parameter values is basically indistinguishable from the surrogate model (on the scale of the plots). We could add these values to the plot, but it would decrease the clarity of the plot, as we would have multiple markers overlapping each other on the plot.
    We have added a clarification in the caption of figure 9 and figure 12.
  - RC2: When describing the "ingredients" of the model, it could be nice to precise which parts are obtained using library calls to UQLab (mentioning the function name of this library could also be interesting to some readers), and which part were implemented in this study.
  - AC2: Our framework UQ4WIND is indeed built using the UQLab toolbox, which contains the in-built algorithms to perform both: sensitivity analysis and Bayesian calibration. We were already giving some details regarding the implementation but will elaborate on this in appendix A, indicating the main function calls. Please note also that the UQ4WIND code is open-source available via the GitHub link in the manuscript. We have added more information on specific UQLab implementation commands in Appendix A.
  - RC3: Is the selection of the PSD peaks a manual process or is it automated? I'm guessing it could be challenging to automate without some kind a knowledge of the system (for instance to distinguish a peak in the low frequency content with potential noise there). Also, the peaks could potentially change with operating conditions/rotational speed. Could you comment a bit on that?
  - AC3: The selection of the PSD peaks is automated. We order the Fourier coefficients in terms of the largest power spectral density and then keep the first few terms. In our case the peaks are easily distinguishable from the noise, and the signals are well represented in terms of a few Fourier coefficients. In case the peaks are close to the noise, a good strategy would be to warn the user and display a plot of the spectrum with the peaks that are to be selected. The reviewer is right that the peaks (position/magnitude) will change with operating conditions, but the Fourier decomposition in terms of amplitude and phase shift takes this into account. Furthermore, in our tests, we typically also visualize the output of the model runs (like in Figure 5 in the

manuscript) and check the match between model output and Fourier representation.
We have added a clarification at the end of Section 3.3.2.

- RC4: Is there a limitation by assuming zero mean here? Could there not be an offset in the quantities of interest, due to some kind of systematic error/ measurement bias? It seems to make sense to have it at zero, but could you justify it briefly?
- AC4: For the sake of simplicity, and also due to the lack of knowledge of the model bias term, the discrepancy term has a zero mean. This is a commonly used approach in Bayesian model calibration, but indeed the reviewer is right that more advanced approaches are possible (e.g. using a Gaussian process to model the discrepancy). We will further clarify this choice and state this explicitly in our revised manuscript. Note, by the way, that the Gaussian discrepancy distribution is one of the predefined likelihood options in the current version of UQLab, but the user can also provide a user-defined likelihood function if available.
  We have added an explanation in Section 4.3, below equation (19).
- RC5: I'm guessing these are different runs than noes used to setup the PC model? Can you precise this?
- AC5: Our apologies for the confusion, N=32 does in fact refer to the number of Aero-Module runs that are used to set-up the polynomial surrogate model, which is then used in the calibration process. This number might seem rather low at first sight, but this is justified by the fast convergence of the LOO error, as explained in appendix B1. The reviewer is right in the sense that these runs are not per se exactly the same as the ones used for the sensitivity analysis, since the sampling method is random (LHS), and the surrogate model of the sensitivity analysis is currently not re-used for the calibration run (since certain parameters are left out based on the sensitivity analysis).
  We have added a clarification at the start of Section 5.1.2.
- RC6: Shouldn't this error increase with the radial position since the loads increase with radius? Could you comment on this?
- AC6: Thank you for this suggestion, this is certainly a possibility. In the current test case, we did not want to introduce too much (possibly wrong or biased) a priori knowledge about the radial dependence, and kept the prior uniform and the same for all radial sections. We then let the calibration process 'do the job' and found indeed higher values of θ_E at the outboard sections than at the inboard sections (see Table 2).
  We have added this insight in the manuscript in section 5.1.2.
- RC7: Could this plot also include the evaluation from the PC model? I'm guessing they would be on top of the calibrated Aero-Module. But I was confused at first when looking at the plot, not knowing if the "Calibrated Aero-Module" was really the Aero-Module, or the PC model. Some precision might help other readers, and I would think having both is quite important.
- AC7: Please see AC1.
- RC8: Can you mention some applications as examples here?
- AC8: The developed framework: UQ4WIND has already found its application for calibrating a dynamic wind farm control model; this is part of our upcoming work. Another topic within wind energy that could benefit from the UQ4WIND framework could be: calibration of low-order acoustic models using empirical correction factors for

wind turbine noise estimation. Furthermore, calibration of engineering wake models, which typically contain several uncertain model parameters (such as wake expansion coefficients), would benefit from calibration using high-fidelity models such as CFD results.
We have added this to the conclusions.