Review of "Hyperparameter tuning framework for calibrating analytical wake models using SCADA data of an offshore wind farm applied on FLORIS"

General comments

The authors present a scheme to calibrate parameters of an analytical wake model. Their method compares power predictions of the model to the power reported in the SCADA data of wind turbines and adjusts the model parameters in three steps to achieve optimal agreement. For validation, the method is applied to an offshore wind farm tuning an analytical wake model. The calibrated model parameters are then analyzed in terms of sensitivity and dependency to wind direction and wind speed.

The overall goal of the paper is to establish a new model calibration method, which is a clear and relevant agenda. However, achieving this goal is currently obstructed by issues of the specific implementation presented here and a missing comparison to other calibration methods for model parameters. In addition, there are some inaccuracies in the paper. I list my most important comments below:

(1) The model parameters are tuned with a cost function between the power predictions of the FLORIS model framework and the power reported in the SCADA data of wind turbines. However, only one part of the full FLORIS model framework, namely the velocity deficit model, has been included in the calibration. This might lead to unrealistic tuning results for the model parameters, if any of the other parts of FLORIS framework not included in the calibration is not set up optimal.
I want to illustrate the above with a specific example in the following. The methods describe the velocity deficit model, but do not provide any information how the velocity field is then related to the power. Specifically, the following points are unclear:

- Which power coefficient and thrust coefficient curves have been used and are they realistic for the wind turbines at the test site?
- If power curves generated from the SCADA data have been used, it would be important to know if they were applied to the model using a rotor averaged wind speed or the wind speed at the nacelle location? Were rotor blockage effects included or not?

The results of the tuning find that the wake growth rate (k_a) has an optimal value that is much lower than typical literature value. However, this outcome might also be caused by a too high power coefficient set in the model framework, which the optimization then tries to correct by reducing the wake recovery.
Therefore, the found differences between the calibrated model parameters and literature values cannot be attributed to better tuning with certainty. While this does not invalidate the tuning method proposed by the authors per se, it is a problem for its validation in my opinion.

(2) Only results achieved with the here proposed calibration method are shown. It would be interesting to include results using model parameters obtained with other calibration methods and standard values from literature. This could be followed by a discussion of the differences between methods.

(3) The discussion of the results refer to many site-specific effects like farm-to-farm interaction and the influence of a nearby shoreline. However, the description of the test site is very sparse for such a discussion.

Specific comments

Line 43-45: This statement should be supported by a citation.

Line 45-46: It is true that measurement errors affect the characterization of the flow state, but the statement seems to be out of place at this point in the manuscript.

Line 66-75: Maybe state the history and parameters for the first model and then for the second model instead of going back and forth throughout the paragraph.

Line 90-92: The paper should motivate the proposed calibration method by pointing out benefits and differences to the already existing calibration approaches referenced here.

Line 108: It should be elaborated what stochastic uncertainty means here. Does it refer to the stochastic error of a mean wind speed that is computed from a finite ensemble of measurements of the turbulent flow? Further possible error sources that can affect it are a drift of the mean value due to diurnal cycle or changing weather conditions. In addition, the mean wind speed can have spatial variation and a single value might not be representative for a large wind farm.

Line 109-111: A wind vane measures the wind direction, but not the wind speed.

Line 130: Referring back to the comment on line 108, turbulence and noise can be adequately quantified with the variance. Other effects like a trend can lead to a non-Gaussian distribution of the measurement values. Would it be possible to extend the proposed framework with other metrics in principle?

Line 164: Does wake blockage refer to farm-to-farm interaction from neighboring wind farms or to wake effects of individual wind turbines within the wind farm?

Section 2.1 in general: The introduction of the test site does not provide sufficient information. Specifically, information should be provided on the topography of the nearby shore, the distance of the wind farm from the shore, distances to neighboring wind farms. The information provided in Figure 2 could extended with precise angles of the affected sectors and distances. Can you provide the location and a map of the wind farm and the surrounding area (if not due to NDA restrictions, that should be stated)?

Figure 3: I assume this figure only shows the filtering described in Section 2.1, but not the steps of Section 2.3 and 2.4. Is it possible to illustrate the impact of the other steps on the data? And should the values for above 25 m/s not be removed as well, because the wind turbines seem to be derating? In addition, the label on the ordinate should be normalized active power.

Line 249-250: How does the ratio of wind speed variance to wind direction variance relate to curtailment of the wind farm? To me, this not clear intuitively and should be further elaborated. I would expect that a low active power for a given wind speed is more indicative of curtailment (in the absence of status alerts or strong wake effects).

Line 257: Cite examples for the use of the model in literature here.

Section 3.1: The choice of using the term GCH might be confusing, because the Curl model outlined in Martínez-Tossas et al. (2019) is not applied here and the Gauss-legacy model is used instead. How is other literature handling the nomenclature?

Line 278-280: Can the other model components affect the results of the parameter tuning of the velocity deficit model presented here? A description of them is missing entirely.

Table 1: Abbreviation SOSFS not introduced. A column could be added to the table providing references to papers describing each of the model component.

Section 3.2: Currently, this part of the paper seems to be separated from the proposed parameter-tuning scheme. How does it contribute to the tuning of the parameters? Should users of the tuning method first conduct a sensitive analysis and remove low sensitivity parameters prior to the model calibration?

Figure 295-298: The parameters k_a and k_b have higher values for wind direction sectors (0, 90) and (180, 270) compared to other wind directions, while it is inverse for alpha and beta. Is there any explanation why that might be the case?

Line 325: What is the benefit of letting the algorithm choose the weighting? It would be problematic if the algorithm chooses a=0.001 and b=0.999 for example and only account for the global power production of the wind farm while the power for individual wind turbines are completely off. The results section should show what values the algorithm chose.

Line 329: What does "freeflow wind turbines" mean? Is it first turbine row of the wind farm at the upstream edge?

Line 366-371: Two questions on the interpretation here:
(1) How does the internal arrangement of the wind turbines inside the wind farm differ for the two clusters presented in Fig. 15 and Fig. 16?
(2) There does not seem to be any effect of shoreline to the south and south-east, which one might expect to also cause problems similar to a neighboring wind farm. Is it further away or is its effect on the incoming free flow less pronounced?
An overview map would be helpful to follow the interpretation of the results here. If a NDA is preventing to include it explicitly, maybe a schematic overview can be provided instead. Plotting the data as a function of the distance to the next heterogeneity upstream might another approach to avoid NDA.

Line 389-391: Isn't it the other way around? The parameter k_a is not inverse to the wake recovery, because the larger k_a is the larger will be the wake recovery for a given turbulence intensity.

Section 4.3 in general: As mentioned in the first general comment, I am not convinced that the tuned parameters are necessarily more realistic values, because other parts of the full FLORIS framework might interfere with the optimization. If this comment cannot be refuted directly, it might be addressed by restricting the validation to a simpler configuration (e.g. running it directly on the wind speed instead of the power (removes the power coefficient from the validation) and using first and second row turbines only (removes wake superposition from the validation)). Another approach might be to assess the sensitivity of the tuned parameter to changes to those other parts of the FLORIS framework.

Line 403-404: There is a contradiction here. The text states that k_a has higher values for the south-east where the coast is, but in Fig. 19 the k_a values for the wind direction sector (180, 250) are lower compared to the sector (50, 180).

Line 404-405: The hypothesis of an increase of k_a with the turbulence intensity could be tested with a plot similar to Fig. 18 and Fig. 19, but with TI on the abscissa.

Captions of most figure can be improved to explain what the figure is showing without having to refer back to the text.

Technical comments

Line 54: Maybe rephrase instead of "no velocity profile" that the Jensen wake model assumes a constant velocity across a wake cross-section.

Line 56-57: Either "by literature (Barthelmie et al., 2009)" or "by Barthelmie et al. (2009)."

Line 123-124: Insert "are" in "effects becoming".

Line 133: Remove "significant".

Line 189-191: Sentence structure not correct.

Line 200: Remove "different".

Line 201: Mean values and variances are not measured directly, but calculated for the 10-minute intervals.

Line 202: Wind turbine should be plural here.

There are several instances where it could considered to combine separate figures into a single figure with multiple panels.