Review of manuscript: WES-2021-152-R1
Title: Multifidelity multiobjective optimization for wake steering strategies
Authors: Quick, King, Barter & Hamlington

<u>Overall comments:</u>

Thank you to the authors for considering the reviewers' comments. The multiobjective optimization methodology proposed by the authors is without doubt a useful, interesting contribution to the wake steering literature. However, I believe the presentation and the interpretation of the results could still be improved. I would like the authors to consider the below comments, with the hope that they can help improve the presentation and the impact of the manuscript's contributions.

<u>Point comments:</u>

1. Page 2, Line 26: "Damiani et al. (2018) performed a detailed analysis of a single wind turbine, noting that negative yaw offsets tended to increase fatigue loading more than positive yaw offsets (although it should be noted that these results were specific to the turbulence seeds used in the study)."
   The added parenthetical remark does not provide the necessary caveat. Damiani et al. (2018) specifically cautioned that the influence of the incident conditions "make generalization more difficult." The authors should state that Damiani et al. (2018) cautioned against the general statement that "negative yaw offsets tended to increase fatigue loading more than positive yaw offsets."

2. Page 2, Line 37: I am still concerned with the phrasing "remarkably accurate in power prediction." This seems to highlight that wake modeling for power predictions is a concluded matter. I don't believe this is the case, as wake models can exhibit high predictive error for many utility-scale wind farm applications, e.g. time-varying conditions, the presence of complex terrain, strong stratification, etc., and this statement is counter to the community calls for research [e.g. 1, 2].

3. Page 3, Line 73: "We use the scikit-learn Gaussian Process implementation (Pedregosa et al., 2011), which is a well-validated open-source project."
   It's useful that the authors have added this, but specifically, I would like to see: 1) a comparison of the GP fit to the training data; 2) a GP prediction (out of sample of the training data) compared to the simulated power and loads for an out of sample set of the design variables.

4. Page 5, Line 118: Many readers of *Wind Energy Science* will not be familiar with Pareto dominance, and this article should be self-contained. Please add one or two sentences defining Pareto dominance before referring to the paper or remove its use from the manuscript.

5. Page 9, Line 221: "These time parameters were justified by comparing power and loading computed over time intervals of 600-900 s and 900-1,200 s, resulting in relative differences of

only 2.6% for power and 4.2% for loading when \gamma = (15;0)."
These differences seem nontrivial. Have the authors tested how sensitive the Pareto set is to these finite-time averages? Does the optimal yaw change 0.1 degrees or 10 degrees? Especially given how sensitive the flow physics interpretation is to the particular wind conditions, I am interested in hearing more about this.

6. Eq. (32): Now that you have clarified that the choice of the objective function (especially the -10 in the loads function) was *ad hoc* to have a negative value, I am wondering about the effect of your objective function here. Is the particular quantity of the objective function entirely arbitrary in your framework or will it impact your Pareto front and/or your exploration/exploitation tradeoff? Perhaps the mean of the objective does not matter but it is the sensitivity with respect to the design variables which matters? Simply stated, a reader will want to know: how do I pick an objective function if I want to apply this proposed method? Should it depend on the turbine model, farm geometry, etc.?

7. It is reasonable to ask whether the presented results actually show that the multifidelity approach is better than the single fidelity. For minimizing EHVI, it seems that multifidelity is better (although the noise is larger for multifidelity and if you extrapolate the trend lines it would seem to suggest the blue line will drop below the orange with more evaluations).
For the end objectives of power and loads, the multifidelity is only faster to estimate minimum loads, it is actually slower for estimating maximum power. One also has to recognize that these are the results where the authors have intentionally created an artificial loads model such that the multifidelity is better than the single fidelity (Page 11, Line 286). Does this bias the results?
I would appreciate for the authors to confront this more directly. The paper is clearly a novel, significant contribution already based on the multiobjective optimization approach. But can it be concluded that multifidelity is superior to single fidelity? Perhaps this is a call to action for better loads modeling more than anything.

References

[1] Veers, Paul, Katherine Dykes, Eric Lantz, Stephan Barth, Carlo L. Bottasso, Ola Carlson, Andrew Clifton et al. "Grand challenges in the science of wind energy." *Science* 366, no. 6464 (2019): eaau2027.
[2] Meneveau, Charles. "Big wind power: seven questions for turbulence research." *Journal of Turbulence* 20, no. 1 (2019): 2-20.