Response to Referee 3

We greatly appreciate the time taken by the referee to read our manuscript. We have taken into consideration and addressed all comments, questions, and suggestions from the reviewer, and we feel that the revised manuscript is now substantially stronger as a result. Changes made to the text at the request of the reviewer have been highlighted in blue in the revised manuscript. In the following, reviewer comments are repeated in italics and our responses are provided in the bulleted sections of text.

General comments
This manuscript proposes the application of the stochastic gradient descent (SGD) method from deep learning to the wind farm optimization problem, which maximizes annual energy production (AEP) with respect to wind turbine location. It is claimed that this work is the first such application, and it shows results demonstrating significant reduction in computation time for a small set of test cases. The manuscript is generally well-written, and the writing is concise.

Specific comments
Results are presented for three wind farms of different sizes, with 20 optimizations each with randomized initial designs. The authors are also asked to consider different wind farm shapes. Since there are some tuning parameters in the algorithm (and some tuning was required to achieve favorable results), a concern is whether significant problem-specific tuning is required. Demonstrating consistent results with different wind farm shapes (beyond just the one rectangular domain) would be very compelling.

- We added material to the manuscript showing the results of circular wind farm layout optimization in addition to examining square wind farms. These changes are reflected in the Methodology on P3L70-80, P5L117, and P8L181, the Application on P8L194-197, as well as a dedicated "circular wind farms" section in the Results.

- As noted in the text on P13L283-P14287, wind farms with circular boundaries yields similar trends in the optimization results to the wind farms with square boundaries.

There is clearly a cost to the reduction in computation time since constraints are not exactly satisfied. Could the authors comment on the degree of constraint violation and its significance, considering a range of different problems? Perhaps it would be helpful to visualize a Pareto front of computation time versus constraint violation, varying η_T.

- We have added a Sensitivity Analysis section to the results to have a dedicated space to answer these and related questions.

- We analyzed the final constraint violations associated with different initial and final learning rates. The analysis indicated that η_T is strongly related to the total constraint violation, as illustrated in the left panel of Figure 11 and the accompanying text in P17L323-P18L329.

- We also performed an analysis of the trade-offs between AEP and constraint satisfaction by changing the initial and final learning rates. We include this in the Sensitivity analysis section, in the right panel of Figure 11, as well as with corresponding discussion on P17L329-333.

- Additionally, we included material showcasing an "early stopping" capability of the optimizer, where AEP computations are neglected near the end of the optimization algorithm. This often

yields solutions that exactly satisfy the imposed constraints. The results are shown in Figure 8 and the accompanying discussion is on P16L294-305. For completeness, this "early stopping" option was explicitly added to Algorithm 1.

Technical corrections

1) *Line 28-29: Please rephrase - sentence is confusing*

- We changed this sentence to enhance the clarity (P2L29-30):
  The algorithm samples the gradient of a stochastic objective, following the mean gradient by a specified distance, then repeating the process, which amounts to optimizing the expected value of the objective.

2) *Line 44 and onwards: "SciPy SLSQP" should be changed to "SLSQP" - SciPy is a just a package that wraps an implementation of SLSQP*

- We have removed the association of Scipy with the SLSQP algorithm from this point and onward.

3) *Line 77: for completeness, it should be stated at this point where the 8760 comes from, or at least mention that the units are hours*

- We have added a note that this is the number of hours in a year on P4L84-85.

4) *Line 107: I suggest separating the method from the implementation. For example, here, the choice to use the finite-difference method and algorithmic differentiation are aspects specific to the implementation, not aspects of the general method*

- We have moved this information about gradient computation to the Application Section on P9L217-218.

5) *Line 128: "intial" - typo*

- We corrected the typo.

6) *Figure 2: why do these curves have these shapes? Please comment in the manuscript on the trends and whether they agree with your intuition*

- We did this analysis again using the TOPFARM and supercomputer setup associated with the revised results (Figure 3).
- We note that the computational cost grows logorithmically with the number of wind turbines and that there is a more complex scaling associated with increasing the number of samples, which may have to do with the way memory slows down the computer (P10L221-230).

7) *Line 200: again, please comment on *why* the gains with SGD are higher with wind farms with more wind turbines*

- In the original manuscript, SLSQP tended to reach the maximum number of iterations as the number of turbines increased. This is reflected in Figure 3 of the originally submitted manuscript, where the deterministic results tend to end at different times when examining small wind farms and the same times when examining large wind farms. So, it's not exactly that SGD did better as the number of wind turbines increased, but rather that SLSQP did worse.

- In our revised results, SGD consistently produced 0.3-0.5% more AEP than the SLSQP counterpart, with the vast majority of results being closer to 0.5% improvement. We discuss this on P13L273-275, remarking that this is likely because the SGD algorithm is able to better explore the design space by initially relaxing the constraints, allowing for some initial constraint violations.

- We have revised the manuscript to reflect that SLSQP generally produces slightly less AEP than SGD, regardless of the wind farm size, and that SGD becomes dramatically less time-consuming to run as the number of turbines is increased (P11L249-P12261 and P13L278-P14291).

8) *Line 214: "constrain" - typo*

- We corrected this typo and ensured this does not happen elsewhere in the text.