This paper compares the performance of multiple AEP models in wind farm layout optimisation. It captures a good subset of the universe of models, and compares them based on their ability to predict AEP, to optimise layouts that yield high AEP, and their execution time, paired with an optimisation algorithm and a stopping criterion.
The experiment is well designed, captures different development areas, turbine counts, and wind rose modalities, so I feel its results can be generalised beyond these scenarios.
This is a good guidance for wind farm layout design, and a good foundation for further studies.
The experiment shown here is what most wind farm designers have to do anyway, and now they won't have to reinvent the wheel and can cite these results instead.
I have minor comments to improve the quality of the text and the discussion.
I recommend the paper is published with minor corrections.
I particularly appreciate that the authors shared the entire code to recreate the results.

Typos

line 151, says "method" and it should say "methods" in plural
line 171, says "third approach" and it should say "second approach"
line 187, a period is missing at the end of the paragraph.
line 193, says "include rotor treated as", should say "include the rotor being treated as"
line 259, says "wind conditions used to ... are used based on", should just say "... are based on"
line 359 says "off-shore" and we normally just say offshore.
line 465 says "considere" instead of "considered"
Line 495, not common to say "took shorter", I'd say "took a shorter time", or "took less time".
Line 500, duplicated "for for"
Line 500, says "results" instead of "result"
Line 504, says "proven" instead of "proved"

General comments

First sentence in the introduction reads "wind turbines are placed in groups known as wind farms", which I think it underestimates the reader of a wind energy journal. I also found it unnecessary to explain in 2.6 what convex and concave polygons are, too basic a concept in geometry to have to explain to a reader.

Section 3.1, which layouts were used to evaluate AEP, random? I think gridded layouts would add information, as coarser models would not necessarily evaluate problematic alignments.

Line 380 says "a model may be suitable to evaluate AEP, but not be useful for WFLO". I understood later in the paper why that is, but at first sight, this could be explained better here.

What is the optimisation stopping criterion? Convergence of AEP, of layout, something else?

Lines 471-473. I believe that the shape of the development area is ignored in the discussion here. You can sample wind directions with BQ, OWQ that do not take into account the geometry of the site, which the optimisation is constrained by.

Line 484, why is the convexity of the boundaries a factor in the performance of the wake model (in your discussions), instead of only being a factor in the performance of the optimisation algorithm? I fail to see the relation.

Line 502, why does BPA-FLOWERS take so long to optimise Sofia?

Line 523, you offload the exploration of the effect of number of WD bins on the model, and I argue that if there's any paper that should explore that, it should be this one. Especially since you write appendices for number of CPUs and number of multistarts.

If you ignored the minimum inter-turbine spacing constraint because it adds complexity and execution time, makes me wonder: in real application we always need to define some sort of proxy for turbine loads, so if they were brought back into the study, would that render your AEP model comparison meaningless? Would the outcome be, "time is dominated by constraints, so just use the reference quadrature model"?

In the conclusion, the paragraph at line 587 repeats much of the previous discussion, like a summary of the conclusions. I suggest removing to avoid redundancy.