The manuscript under review discusses wind farm layout optimization, considering both energy production and the cost of the electrical network. The former is based on evaluating wind turbine farm layouts with a fast Gaussian wake model. The latter is based on a simple graph-theoretic approach where straight lines between turbines are assumed and a small number of different cable cross sections are available, with different capacities and costs. In particular, neither the topography (e.g. necessary burial depth) nor its dynamics (e.g. traveling sand waves) are considered.

Optimization is either performed sequentially, where first the power production is optimized and then the layout, or simultaneously, where first the total cost is optimized, using a simplified ("heuristic") model for the electrical layout, and then the layout is again optimized. The latter is necessary, since for the heuristic model two of four constraints mentioned are relaxed. In both approaches the optimal layout in the final step is calculated with a mixed integer linear program (MILP) that respects all constraints.

The topic is relevant and the manuscript is mostly well written. However, there are slightly too many formulas (the authors use formulas even in cases where some text would suffice), and some notation that occurs is not explained (see below for details) - this should be reviewed and improved.

The authors have a good point in that there is a major tradeoff between optimization for production (widely spaced wind turbines) and optimization for cable layout cost (closely spaced turbines) that has potential for further optimization compared to the sequential approach, and their results seems to demonstrate this, with an up to 6 percent improvement "on the quality of fully feasible wind farm designs" using their simultaneous approach. This is promising.

My main technical question is actually on the (previously published) formulation of the cable layout optimization problem as a MILP. Is this not much too complicated? Why is it not sufficient to only use the connectivity variables $x_{ij}$ and optimize their values, under suitable constraints? It seems that the objective function and all constraints can be expressed as functions of the $x_{ij}$. Or am I mistaken? And what is the role of the $f()$ function? Why are the $y_{ij}^{k}$ variables needed? What is the "maximum number of WTs connectable through an arc"? The MILP formulation has to be better explained!
Major comments

- Optimization of wind turbine positions is either using no cable layout cost, or a fast, heuristic approach. Is it not possible to run the mixed integer linear programming (MILP) cable optimization during this optimization? Is it so slow? This would provide reference results against which the results of both approximations could be compared to.
- The constraint C1 enforcing a tree topology: Why are no back-connections considered, offering a redundant electrical path in case of a fault at one turbine? Of course this would complicate the optimization problem and its formulation, but why are these technical solutions excluded a priori? The layout cost would be higher, of course, but OPEX or financial risk might be lower due to the redundancy.
- page 7: Implementation of constraints. The approach only works since the methods in Step 1 are all greedy algorithms, right? And even then there are problems, as the authors mention (infeasible points can be obtained), unless a better approach is found. Simply relaxing the constraints (not considering C3 and C4, for example) is a somewhat radical solution. But since this has been done, it would be interesting to know how often the cable layout was infeasible before running the MILP for it.
- Algorithm 1: needs comments! Otherwise this is hard to read. And what is the inverse of an arc and why is it needed? (Why are directed graphs considered here, not undirected ones?) What are its inputs? What idea does the tradeoff value T represent intuitively?
- page 7: Unclear notation. What is the set "B" in the definition of the "trivial optimization problem"? And what is "X_c"?
- Fig. 6: this algorithm could be explained better, e.g. with fewer formulas
- What about uncertainty (e.g. in the spot price) - how would that influence the optimization results?
- Table 3: I assume that U is given in terms of number of turbines supported by the respective cable? Or what are its units?
- page 16: Unclear notation. "correspond to t_r \approx 0" - What is t_r?
- Unclear notation: What is IRR_i and what is IRR_0?
- Table 4: Do not use komma, but decimal point. Too many decimals given. Why are results for both one-tailed and two-tailed tests given?
- Conclusions: "an improvement of the IRR of 3.52% percent is achieved..." - This is when comparing the sequential and simultaneous heuristic approaches, I assume? More interesting would be, as already mentioned above, how large the difference with the optimum (as derived by MILP) would be between the heuristic-derived and the MILP layout.
- "Code available upon request": Nowadays, statements like that are hard to justify, as it is so easy to upload the code to an institutional or other repository (e.g. Github).
- Data available upon request: This is marginally acceptable, but the data should ideally be uploaded into a public repository (e.g. Zenodo).

Minor comments

- page 5: "wind directional bits" should probably be "... bins"?
- page 21: "with higher quality than the simultaneous counterpart" - this should probably be "... than the sequential counterpart"?