

Dear Michael Muskulus,

Thanks for taking your time to review our manuscript, for suggesting valuable changes for it, and for appreciating the contributions of our research. See changes after your comments in the new version of the article in Purple color. Find below responses to each of your comments:

**COMMENT 1:** 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!

**RESPONSE 1:**  $y_{ij}^k$  is the main variable in the model. This variable models the decision of how many  $k$  wind turbines are connected by the arc  $(i,j)$ .  $x_{ij}$  is an auxiliary variable that allows for a more intuitive and compact full representation of the model, as it represents which arc  $(i,j)$  is active. Our hypothesis is that as both variables are linked through Equation (8), this may help tighten up the model and work as valid inequalities. The objective function therefore has to be expressed in function of variables  $y_{ij}^k$ , which encode the cost value of the optimum cable to support those  $k$  wind turbines, computed in a preprocessing stage. The function  $f()$  allows reduction of variables number as it computes the maximum number of wind turbines connectable through arc  $(i,j)$ , so for example, only arcs that are connecting directly to the OSS can support the maximum number of WTs  $U$ , which is not true for arcs connecting to a wind turbine supporting maximum  $U - 1$  turbines.

Some sentences have been added to the model description as presented in lines 215 to 220, as follows:

Eq. (2) defines the variables of the model. Binary variable  $x_{ij}$  is one if the arc  $(i,j)$  is selected in the solution, and zero otherwise. Likewise, binary variable  $y_{ij}^k$  models the  $k$  number of WTs supported upstream (with respect to flow towards the OSS) from  $j$ , including the WT at node  $j$  (under the condition that  $x_{ij} = 1$ ). Function  $f(i)$ , that allows for variables number reduction, maps from tail node  $i$  to maximum number of WTs connectable  $k_m$  through an arc  $(i,j)$ . If  $i = 1$  (i.e. the OSS), then  $k_m = f(i) = U$ , otherwise  $k_m = f(i) = U - 1$ . Eq. (3) is the objective function. Cost parameter  $c_{kij}$  encodes the optimum cost to connect  $k$  WTs through arc  $(i,j)$  and is obtained similarly to Step 2 in Fig. 2. [C2] is enforced at this point as well. Eq. (4) ensures simultaneously a tree topology, only one cable type used per arc, and the head (j)-tail (i) convention, while Eq. (5) is the flow conservation which avoids forest graph; both Eq. (4) and Eq. (5) guarantee [C1]. Eq. (6) expresses the requirement of [C3]. The set  $\chi$  stores pairs of arcs  $\{(i,j),(u,v)\}$ , which are crossing each other. Excluding crossing arcs ([C4]) in the solution is ensured by the simultaneous application of Eq. (7) and Eq. (8). Finally, Eq. (9) defines a set of valid inequalities to tighten the mathematical model.

**COMMENT 2:** 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.

**RESPONSE 2:** As stated in line 355 of the manuscript, the exact solver for the MILP model can take up to 10 hours and no less than dozens of minutes (in rare cases) to terminate given a minimum gap of 2%. Therefore, it is unpractical to run this solver during the simultaneous optimization. We rely on the presented heuristics as a proxy to estimate the cable layout cost.

**COMMENT 3:** 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.

**RESPONSE 3:** This manuscript addresses the collection system problem in OWFs considering only radial designs. First, this is because a majority of OWFs developers only consider networks with this topology type. Second, cable loops can be studied with accentuated focus in a subsequent stage, where after reliability analysis, redundant cables can be installed in the radial system, in order to account for contingencies of cables and wind turbines. In the operating mode, OWFs generally maintain the switchgears of those redundant cables in open circuit, maneuvering these components after a fault in an electrical component has been identified. Finally, we have conducted studies regarding reliability-based optimization (see article “Reliability-based Topology Optimization for Offshore Wind Farm Collection System”, <https://onlinelibrary.wiley.com/doi/full/10.1002/we.2660>). Results indicate that the benefits to design redundant topologies depend on parameters such as Mean Time Between Failures (MTBF), and Mean Time To Repair (MTTR), with rather limited set of reasonable available data.

**COMMENT 4:** 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.

**RESPONSE 4:** The presented greedy heuristics are very fast, however they have issues of scalability in terms of supporting a large number of constraints. We have chosen to respect constraints C1 and C2, as they are rather straight-forward to be included by the heuristics. The benefits of relaxing constraints C3 and C4 are twofold. First, computing time is saved trying to satisfy them, process that can be cumbersome (see article of the first author, <https://arxiv.org/abs/2108.13973>). Second, as discussed in the manuscript, by relaxing the constraints a counteracting effect is theoretically achieved, as heuristics inherently provide suboptimal solutions when fully satisfying all constraints. It is out of interested in this manuscript to actually present the satisfiability of the constraints for each WT layout. Instead, the reflection over cost is presented in Figures 9 and 10, where these are compared to the solution of the exact optimizer.

**COMMENT 5:** 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?

**RESPONSE 5:** Thanks for the observation, we have tried to improve the description of Algorithm 1, please see the new comments in lines 180 to 190 of the manuscript. The definition of an arc ( $a = (i, j)$ ) and its inverse ( $\bar{a} = (j, i)$ ) are clarified in the manuscript. Because the directionality of arcs matter, the input must be given in form of a directed graph. The directionality has a clear influence over the trade-off values. The trade-off values represent the balance between the length of an arc and the estimated distance to the OSS; therefore, priority is given to connect WTs located the farthest to the OSS for a given length.

New description:

The input of the algorithm are collected in the weighted directed graph. The first five lines initialize the useful sets. The most important is the trade-off set  $T_o$  using the weight parameters  $p_i$ ; note that due to the weight parameter definition an arc  $a = (i, j)$  and its inverse  $\bar{a} = (j, i)$  must be considered independently. The iteration process starts at line six and continues until a fully connected tree graph is obtained. In each iteration the arc with the lowest trade-off value is incorporated to the tree, as long as it satisfies [C1] and [C2] (lines seven to nine). According to the trade-off value, priority is given to the arcs located the farthest to the OSS for the same arc length, as the greater  $p_i$ , the lower  $t_a$ . In case constraints are met, the component sets (that include nodes  $i$  and  $j$  of the selected arc  $a_o$ ) are merged, and the trade-off values linked to the newly formed component are updated (lines 10 to 14). Otherwise, the arc  $a_o$  and its inverse  $\bar{a}_o$  become completely banned from the design process, by equalizing their trade-off values to infinite (lines 16 to 17). The indirect graph  $G_d$  in line 19 contains the graph tree nullifying any directionality of arcs.

**COMMENT 6:** page 7: Unclear notation. What is the set "B" in the definition of the "trivial optimization problem"? And what is " $X_c$ "?

**RESPONSE 6:** Thanks for the observation, the text of lines 190 to 195 has been updated to:

Let  $\beta$  be the number of WTs connected through edge  $e \in E_o$  with length  $d_e$ . The following trivial optimization problem must be solved:  $\min \{X^T c \cdot C \cdot d_e : X^T c \cdot U \leq \beta, \|X_c\|_1 = 1, X_c \in B|T\}$ , being  $B|T|$  the  $|T|$ -tuple binary set. The solution provides the cheapest cable  $t \in T$  able to support  $\beta$  WTs via binary variable  $x_{ct} \in X_c$ , where each tuple  $(e \in E_o, \beta)$  defines an independent problem solved in linear running time.

**COMMENT 7:** Fig. 6: this algorithm could be explained better, e.g. with fewer formulas

**RESPONSE 7:** We consider that the formulas help to formalize the operating principles of the algorithm. The text from lines 296 to 303 complement the Fig. 6. Additionally, the reference (Feng and Shen, 2015) could be read for more details of the algorithm.

**COMMENT 8:** What about uncertainty (e.g. in the spot price) - how would that influence the optimization results?

**RESPONSE 8:** The following paragraphs discuss the influence of input parameters variability:

At the end of Section 3:

The presented results are project-dependent, and therefore are vulnerable to variations of the set of input parameters. An increase in energy price would result in a heavier weighting of AEP during the optimization process, giving priority to a more spread out WT layout. The cable costs are another important parameter, that in turn if greater, could favor more strongly the cable layout design by bringing together the WTs; similar logic applies for OSS location further away to the WTs. The available area is also deemed as key, since larger areas would give more room to exploit trade-offs between WT and cable layouts when applying the design Approach 2.

At the end of the conclusions:

Typical values for the input parameters (energy price, cable costs, available area, OSS location, among others) are utilized to set up the case study. However, the outperforming capacity of Approach 2 over Approach 1 is affected by variations of those conditions: more expensive energy prices can result in a better performance of a sequential design approach, while greater cable costs could have the opposite effect. A more detailed analysis of the impact of those parameters over the results of the optimization could constitute future work.

**COMMENT 9:** Table 3: I assume that U is given in terms of number of turbines supported by the respective cable? Or what are its units?

**RESPONSE 9:** Thanks for the observation. Yes, it is the number of wind turbines. This description has been added to Table 3.

**COMMENT 10:** page 16: Unclear notation. "correspond to  $t_r \approx 0$ " - What is  $t_r$ ?

**RESPONSE 10:** Thanks for the observation. This is the computing time of the Random Search algorithm. This has been clarified in the text as:

$t_r \approx 0$  (computing time of the random search algorithm, see Fig. 6)

**COMMENT 11:** Unclear notation: What is  $IRR_i$  and what is  $IRR_0$ ?

**RESPONSE 11:** Both variables are defined in Figure 4 and 5. They are the IRR value before and after the Task 2, respectively.

**COMMENT 12:** 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?

**RESPONSE 12:** Thanks for the observation. Both Tables have been corrected correspondingly.

**COMMENT 13:** 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.

**RESPONSE 13:** We have added the following sentence to the conclusion:

The heuristic algorithm for the cable layout performs reasonably well for layouts arrangements with higher density, with an average absolute relative cost difference with respect to the global optimization model of 8%.

**COMMENT 14:** "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).

**RESPONSE 14:** Thanks for the recommendation. A GitHub repo has been created for the code.

**COMMENT 15:** Data available upon request: This is marginally acceptable, but the data should ideally be uploaded into a public repository (e.g. Zenodo).

**RESPONSE 15:** Thanks for the recommendation. A Zenodo repo has been created for the data.

**COMMENT 16:** page 5: "wind directional bits" should probably be "... bins"?

**RESPONSE 16:** Thanks for noticing the typo. We have fixed it.

**COMMENT 17:** page 21: "with higher quality than the simultaneous counterpart" - this should probably be "... than the sequential counterpart"?

**RESPONSE 17:** Thanks for noticing the typo. We have fixed it.