Optimizing wind farm cable routing considering power losses

Fischetti, Martina; Pisinger, David

Published in:
European Journal of Operational Research

Link to article, DOI:
10.1016/j.ejor.2017.07.061

Publication date:
2018

Document Version
Peer reviewed version

Optimizing wind farm cable routing considering power losses

Martina Fischetti
Vattenfall BA Wind,
and
Technical University of Denmark, Operations Research,
DTU Management Produktionstorvet, 424, DK-2800 Kgs. Lyngby
martina.fischetti@vattenfall.com

David Pisinger
Technical University of Denmark, Operations Research,
DTU Management Produktionstorvet, 424, DK-2800 Kgs. Lyngby
dapi@dtu.dk

Wind energy is the fastest growing source of renewable energy, but as wind farms are getting larger and more remotely located, installation and infrastructure costs are rising. It is estimated that the expenses for electrical infrastructures account for 15-30% of the overall initial costs, hence it is important to optimize their design. This paper focuses on offshore inter-array cable routing optimization. The routing should connect all turbines to one (or more) offshore substation(s) while respecting cable capacities, no-cross restrictions, connection-limits at the substation, and obstacles at the site. The objective is to minimize both the capital that must be spent immediately in cable and installation costs, and the future reduced revenues due to power losses. We present a Mixed-Integer Linear Programming approach to optimize the routing using both exact and math-heuristic methods. In the power losses computation, wind scenarios are handled efficiently as part of the preprocessing, resulting in a model of only slightly larger size. A library of real-life instances is introduced and made publicly available for benchmarking. Computational results on this testbed show the viability of our methods, proving that savings in the order of millions of Euro can be achieved.

Key words: Metaheuristics, Mixed-integer linear programming, Matheuristics, Wind farm optimization, Benchmark instances.

1. Introduction

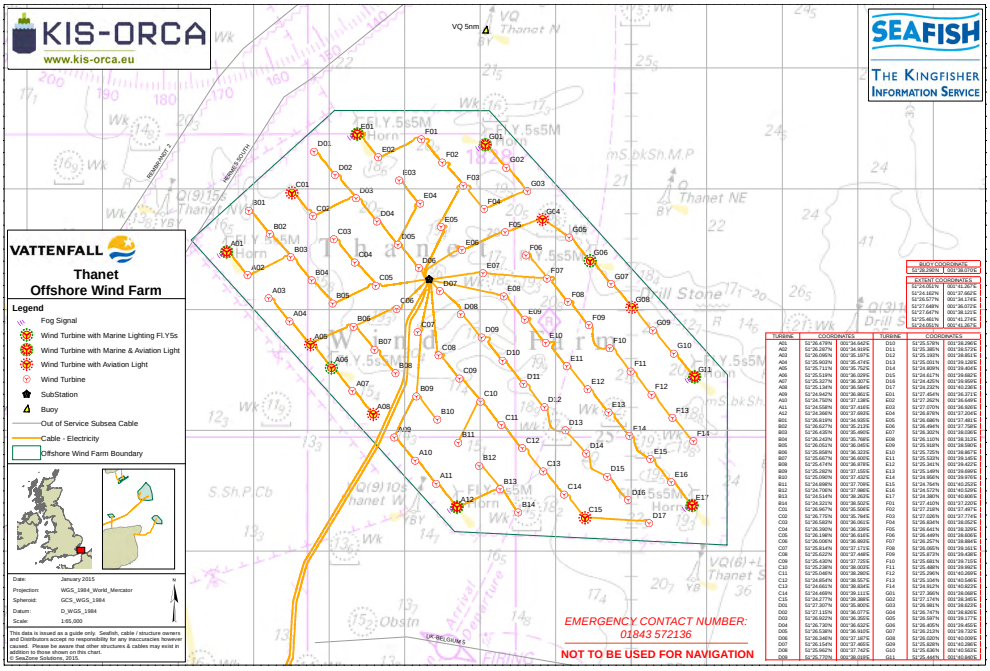
Wind power is an important technology in the transition to renewable energy, fighting climate changes. Designing a wind farm is, however, a complex process including selection of the right site, optimizing the location of each turbine (González et al. 2014, Fischetti and Monaci 2015), establishing the infrastructure (Bauer and Lysgaard 2015) and connecting the farm to the existing electrical grid (Qi et al. 2015). According to González et al. (2014) the expenses for electrical infrastructure account for 15-30% of the overall initial costs of an offshore wind farm. It is therefore

very important to optimize the cable connections among turbines not only from an installation cost perspective, but also considering the power losses during operation.

Thanks to the collaboration with Vattenfall BA Wind it has been possible to build a detailed model including all the constraints arising in practical applications (some of which are missing in previous work from the literature) and to measure, for the first time, the savings in the long run by optimizing the layout and type of cables while taking power losses into account.

The power production of offshore turbines needs to be collected through one or more substations and then conveyed to the coast. To do that, each turbine must be connected through a cable to another turbine, and eventually to a substation. Figure 1 gives an example of cable layout for a real wind park. The final cable layout, often called a *cable routing*, has a tree structure where the non-root nodes correspond to the given turbines, the substations play the role of roots, and the energy (i.e., electric current) flows from the nodes to the roots along the tree.

Figure 1 An example of cable routing for a real-world offshore wind park (Thanet) owned by Vattenfall—image from (Kis-orca 2015).



A number of constraints must be taken into account when designing a feasible cable routing. First of all, the energy flow is unsplittable, i.e., the flow leaving a turbine must be supported by a single cable. In addition, each substation has a physical layout that imposes a maximum number of entering cables.

An important practical constraint is that cable crossings should be avoided. In principle, cable crossing is not impossible, but is strongly discouraged in practice as building one cable on top of

another is more expensive and increases the risk of cable damages. Another restriction is due to the possible presence of obstacles in the site, e.g., nature reserves, already existing wind farms or cables, or terrain irregularities. As a consequence, the final cable routing needs to avoid these areas.

Different types of cable with different costs, capacities and electrical resistances are available on the market. Therefore, one has to optimize also the cable type selection in order to deliver all the energy production to the substations while minimizing both direct costs (i.e., cable and installation costs) and the future revenue losses due to power losses along the cables. This latter aspect is very important in practice, in that more expensive cables/layouts can be more profitable in the long run if they limit the amount of energy lost along the cables. As far as we know, power losses were not addressed in previous work from the optimization literature, perhaps because the loss is a nonlinear function of the current flowing in the cable, hence being more difficult to handle in a Mixed Integer Linear Programming (MILP) framework. Wind park cable routing optimization has obtained considerable attention in the last years. Due to the large number of constraints and the intrinsic complexity of the problem, many studies (i.e., Dutta and Overbye (2011a), González-Longatt and Wall (2012), Li et al. (2008), Zhao et al. (2009)) preferred to use ad-hoc heuristics. Just few articles from the literature use Mixed Integer Linear Programming (MILP) for cable routing; see e.g. Bauer and Lysgaard (2015), Hertz et al. (2012), Fagerfjall (2010), Dutta (2012), Berzan et al. (2011). To the best of our knowledge, only Cerveira et al. (2016) has considered power loss in cables. However, Cerveira et al. (2016) does not take into account variable cable loads due to fluctuating wind. Our work seems to be the only one considering obstacles and power losses over different wind scenarios. An *open vehicle routing* approach was proposed by Bauer and Lysgaard (2015), but their model requires that only one cable can enter a turbine, a condition that is not imposed in our real-world cases. Different solution approaches were proposed in Berzan et al. (2011), where a divide-and-conquer heuristic is proposed together with an Integer Programming model that is tested on small cases involving up to 11 turbines. Hertz et al. (2012) study cable layout for onshore cases. The onshore cable routing problem has however some main differences compared to the offshore one. First of all, cables can be of two types: underground cables (connecting turbines to other turbines or to the above-ground level), and above-ground cables. In the first case, the cables need to be dug in the ground. Due to the fact that parallel lines can use the same dug hole, parallel structures are preferred (up to a fixed number). The above-ground level cables need to follow existing roads. These constraints do not exist in the offshore case. Finally, Dutta and Overbye (2011b) present a clustering heuristic for cable routing.

In this paper we present a new MILP model that is able to handle all the real-world constraints above. Our goal was the design of a practical optimization tool to be used to validate, on real cases, the potential savings resulting from power-loss reduction.

The paper is organized as follows. Section 2 describes our MILP model and is divided into subsections where a basic model is first presented and then improved and extended. In particular, we show how to model power losses, and propose a precomputing strategy that is able to handle this non-linearity in a very simple way, thus avoiding sophisticated quadratic models that would make our approach impractical. We also prove that our specific version of the cable routing problem is NP-hard. Section 3 introduces MILP-based heuristics in the so-called *matheuristic* framework (Boschetti et al. 2009, Hansen et al. 2009). We first propose a relaxation of the initial MILP model that allows one to quickly find an initial (possibly infeasible) solution, and then we introduce different ways of defining a restricted MILP to improve the current-best solution. This approach is combined with an exact solution method to obtain a hybrid heuristic/exact solution method whose performance is investigated in Section 4 on a testbed of real-world cases—input data being available, on request, from the first author. Section 5 analyzes a real case provided by Vattenfall, namely Horns Rev 3, and quantifies the savings obtained by an optimized cable routing taking power losses into account. Some conclusions and future directions of work are finally addressed in Section 6.

2. MILP model

We first introduce the basic MILP model, discuss complexity, and then describe various extensions of the model.

2.1. Basic model

A first step of designing a wind farm layout is to locate the turbines to maximize wind energy capture while minimizing the wake loss. Various techniques, mainly heuristic, have been proposed for this first step (Samorani 2013, González et al. 2014, Kusiak and Song 2010, Archer et al. 2011, Fischetti and Monaci 2015).

Assuming that the best turbine positions have been identified, the next step is to to find an optimal cable connection among all turbines and the given substation(s), minimizing the total cable cost (excluding power losses, as these will be addressed later). Our model is based on the following requirements:

- the energy flow leaving a turbine must be supported by a single cable;
- different cables, with different capacities and costs, are available;
- the energy flow on each connection cannot exceed the capacity of the installed cable;
- a given maximum number of cables, say C, can be connected to each substation;
- cable crossing should be avoided.

Let us consider the turbine positions as the nodes of a complete and loop-free directed graph $G=(V,A)$, and all possible connections between them as directed arcs. Some nodes correspond to the substations that are considered as the roots of the distribution network, and are the only nodes that collect energy. We also add some *Steiner* nodes to add some flexibility to the cable structure, with the additional constraint that at most one cable can enter and exit each of them—hence these nodes can be left uncovered by the cable routing. As explained in Subsection 2.4, these dummy nodes are useful when considering obstacles in the area, or to allow for curvy connections between two nodes.

All nodes $h \in V$ have associated coordinates in the plane, that are used to compute distances between nodes as well as to determine whether two given line segments $[i,j]$ and $[h,k]$ cross each other, where $[a,b]$ denotes the line segment in the plane having nodes $a,b \in V$ as endpoints. In our application, two line segments meeting at one extreme point do not cross. Analogously, two segments do not cross if one is contained in the other, as they represent two parallel cables that can be physically built one besides the other without crossing issues.

We partition the node set V into (V_T, V_0, V_S), where V_T contains the nodes corresponding to the turbines, V_0 contains the nodes corresponding to the substation(s), and V_S contains the Steiner nodes (if any). Furthermore, let $P_h \geq 0$ denote the power production at node $h \in V$, where $P_h > 0$ for $h \in V_T$ and $P_h = 0$ for $h \in V_S$ (P_h being immaterial for $h \in V_0$).

Let T denote the set of different types of cable that can be used. Each cable type $t \in T$ has a given capacity $k_t \geq 0$ and a unit cost $u_t \geq 0$. Arc costs

$$c^t_{i,j} = u_t \cdot dist(i,j)$$

are defined for each arc $(i,j) \in A$ and for each type $t \in T$, where $dist(i,j)$ is the Euclidean distance between nodes i and j.

In our model, for each arc $(i,j) \in A$ we use a continuous variable $f_{i,j} \geq 0$ to represent the (directed) energy flow from i to j, and the binary variable $x^t_{i,j}$ with the following meaning:

$$x^t_{i,j} = \begin{cases} 1 & \text{if arc } (i,j) \text{ is constructed with cable type } t \\ 0 & \text{otherwise.} \end{cases} \qquad (i,j) \in A,\ t \in T.$$

Finally, binary variables $y_{i,j}$ indicate whether an arc (i,j) is built with any type of cable, i.e.,

$$y_{i,j} = \sum_{t \in T} x^t_{ij}, \quad (i,j) \in A.$$

For the sake of generality, in our model we allow the costs $c^t_{i,j}$ to be defined arbitrarily. In addition, we define the undirected edge set $E = \{\{i,j\} : (i,j) \in A\}$ and generalize the non-crossing

property by considering a generic input set $\mathcal{C} \subset E \times E$ of *crossing edges* with the property any two arcs (i,j) and (h,k) cannot be both constructed if $(\{i,j\},\{h,k\}) \in \mathcal{C}$.

Our basic MILP model then reads:

$$\min \sum_{(i,j)\in A} \sum_{t\in T} c^t_{i,j} x^t_{i,j} \tag{1}$$

$$\sum_{t\in T} x^t_{i,j} = y_{i,j}, \quad (i,j) \in A \tag{2}$$

$$\sum_{i\in V: i\neq h} (f_{h,i} - f_{i,h}) = P_h, \quad h \in V_T \cup V_S \tag{3}$$

$$\sum_{t\in T} k_t \; x^t_{i,j} \geq f_{i,j}, \quad (i,j) \in A \tag{4}$$

$$\sum_{j\in V: j\neq h} y_{h,j} = 1 \quad h \in V_T \tag{5}$$

$$\sum_{j\in V: j\neq h} y_{h,j} = 0, \quad h \in V_0 \tag{6}$$

$$\sum_{j\in V: j\neq h} y_{h,j} \leq 1, \quad h \in V_S \tag{7}$$

$$\sum_{i\in V: i\neq h} y_{i,h} \leq 1, \quad h \in V_S \tag{8}$$

$$\sum_{i\in V: i\neq h} y_{i,h} \leq C, \quad h \in V_0 \tag{9}$$

$$y_{i,j} + y_{j,i} + y_{h,k} + y_{k,h} \leq 1, \quad (\{i,j\},\{h,k\}) \in \mathcal{C} \tag{10}$$

$$x^t_{i,j} \in \{0,1\}, \; (i,j) \in A, t \in T \tag{11}$$

$$y_{i,j} \in \{0,1\}, \; (i,j) \in A \tag{12}$$

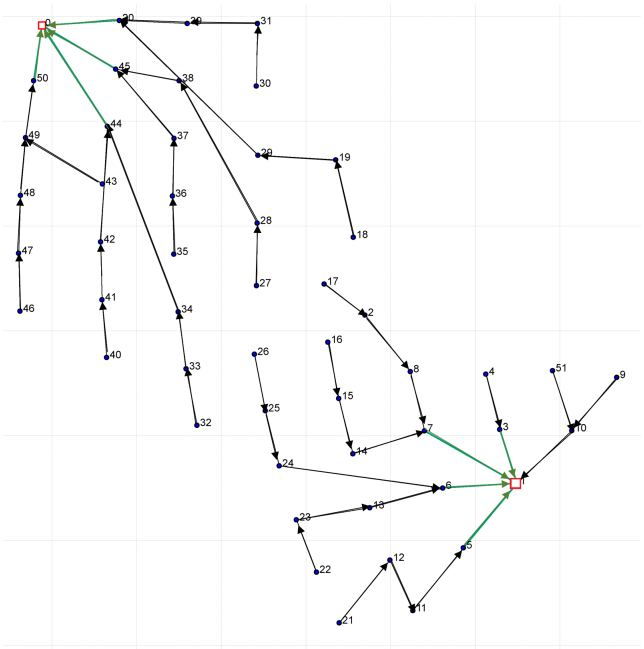
$$f_{i,j} \geq 0, \quad (i,j) \in A. \tag{13}$$

The objective function (1) minimizes the total cable layout cost. Constraints (2) impose that only one type of cable can be selected for each built arc, and define the $y_{i,j}$ variables. Constraints (3) are flow conservation constraints: the energy (flow) exiting each node h is equal to the energy entering h plus the power production of that node. Note that these constraints are not imposed for $h \in V_0$, i.e., when h corresponds to a substation. Constraints (4) ensure that the flow does not exceed the capacity of the installed cable. Constraints (5) impose that only one cable can exit a turbine and constraints (6) that none can exit the substation (tree structure with root in the substation). Note that the Steiner nodes can be connected or not, and if connected only one cable can enter these nodes (constraints (8)). In addition, (7) imposes that only one cable can exit a Steiner node. Constraint (9) imposes the maximum number of cables (C) that can enter each substation. Finally, no-cross constraints (10) forbid building any two incompatible arcs.

Note that the above constraints imply that the y variables define a set of connected components (one for each substation), each component containing a directed tree with out-degree at most one (i.e an anti-arborescence) rooted at a substation; see Figure 2 for an illustration. Circuits are not explicitly forbidden in the model, however because of (3) they can only arise among Steiner nodes (for which $P_h = 0$) and involve at least 3 such nodes because of (10). As explained in the following, this possibility is allowed in our model and used to take obstacles into account.

We finally observe that one could easily extend our model to also include a maximum number of cables entering each turbine (by adding a constraint similar to (9) for each turbine) or to limit the number of cables $\sum_{(i,j)\in A} x^t_{i,j}$ used for each cable type t.

Figure 2 A feasible cable routing; arcs are directed according to the electric current flow, i.e., towards roots (substations).



2.2. Problem complexity

We next address the complexity of the cable routing problem defined in the previous section. The first theorem considers the case where all turbines have the same power production $P_v = 1$ and the nodes are not associated with Euclidean coordinates, the second one considers the case where the turbines are allowed to have different production and they are associated with points on a plane.

THEOREM 1. *The cable routing problem is strongly NP-hard even if* $P_v = 1$ *for all* $v \in V_T$, $|T| = 1$, $C = 2$, $|V_0| = 1$, $V_S = \emptyset$, *and* $\mathcal{C} = \emptyset$.

Proof. We prove the claim by reduction from the Weight Constrained Graph Tree Partition Problem (WGTPP). The latter problem is defined as follows. An undirected graph $G=(V,E)$ is given, with associated cost c_e for each edge $e\in E$, and weight w_v for each node $v\in V$. The problem is to partition the node set V into p disjoint clusters U_r and to build on each of them a spanning tree T_r. The objective is to find a partition such that the overall tree cost $\sum_{r=1}^{p}\sum_{e\in T_r} c_e$ is minimized, while ensuring that each cluster satisfies a weight constraint $\sum_{v\in U_r} w_v \le W$.

In Cordone and Maffioli (2004) the WGTPP problem was proven to be strongly NP-hard by reduction from SAT. Actually WGTPP remains strongly NP-hard even if G is a complete graph, all nodes v have the same weight $w_v=1$, $p=2$, and the edge costs satisfy the triangle inequality.

Given an instance of the WGTPP on a complete graph and with $p=2$ and $w_v=1$ for all $v\in V$, we transform it to a cable routing problem by using the same vertices as turbines having production $P_v=w_v=1$. An extra node is added to represent a single substation, with connection cost $-M$ with all other nodes, where M is a sufficiently large positive value. The substation can be connected to at most $C=p$ cables, and only one cable type exists ($|T|=1$) with capacity $k_1=W$. No Steiner nodes are present ($V_S=\emptyset$), and $\mathcal{C}=\emptyset$. By construction, an optimal solution of the cable routing problem yields the required optimal WGTPP partition.

□

Interestingly for our wind-farm application, our cable routing problem remains NP-hard even in its *2D-Euclidean version*, i.e., when nodes have associated coordinates in the plane, costs depend on the Euclidean distance, and set $\mathcal{C}$ is defined according to the geometrical crossing property between line segments in the plane.

THEOREM 2. *The 2D-Euclidean cable routing problem is NP-hard even when $V_S=\emptyset$, $C=\infty$, and $|T|=1$.*

Proof. We use a reduction from the following well-known NP-complete problem (Garey and Johnson 1979):

> `PARTITION`: Given a set of n positive integers $r_1,\cdots,r_n$, does there exist a set $Q\subset\{1,\cdots,n\}$ such that $\sum_{h\in Q} r_h=\sum_{h=1}^{n} r_h/2$?

Given any instance of `PARTITION`, we define an instance of our cable routing problem as follows. We define $V_0=\{0\}$, $V_T=\{1,\cdots,n\}$, $V_S=\emptyset$, and $P_h=r_h$ for $h=1,\cdots,n$. We define only one type of cable (i.e., $T=\{1\}$) with unit cost $u_1=1$ and capacity $k_1=\sum_{h=1}^{n} r_h/2$, and set $C=\infty$. As to point coordinates in the plane, we position the substation node 0 at coordinates $(0,0)$, while all turbine points $h\in V_T$ are located at coordinates $(1,0)$. As all points lay on a line, no crossing can arise as the line segments corresponding to the arcs of G are parallel. Furthermore note that, as cable cost depends on the cable length (which is zero between overlapping points), it is always more

convenient to connect turbines one to each other than to the substation (until the cable capacity is exceeded). By construction, the answer to `PARTITION` is yes if and only if the optimal cable routing problem has an optimal value of 2, meaning that only 2 cables are connected to the substation, each with saturated capacity. This completes the proof of correctness of our reduction. □

2.3. Improved no-cross constraints

No-cross constraints (10) in our basic model have two main drawbacks: they are weak in polyhedral terms, and their number can be very large (up to $O(|V|^4)$ constraints). We therefore propose a clique strengthening of these constraints, that exploits constraints (5)-(7) to reduce their number and to typically improve their quality. For any node triple (a,b,k), let the *clique* arc subset $\mathcal{Q}(a,b,k)$ be defined

$$\mathcal{Q}(a,b,k) = \{(a,b),(b,a)\} \cup \{(k,h) \in A : (\{a,b\},\{k,h\}) \in \mathcal{C}\} \tag{14}$$

THEOREM 3. *The following* improved no-cross constraints *are valid for model* (1)-(13)*:*

$$\sum_{(i,j)\in\mathcal{Q}(a,b,k)} y_{i,j} \le 1, \quad a,b,k \in V,\ |\{a,b,k\}| = 3 \tag{15}$$

Proof. Let (y,x,f) be any feasible solution of model (1)-(13). As the solution cannot contain 2-node circuits, we have $y_{a,b} + y_{b,a} \le 1$. If $y_{a,b} = y_{b,a} = 0$, the claim follows from the out-degree inequalities (5)–(7) that imply $\sum_{h\in V:h\neq k} y_{k,h} \le 1$. Otherwise, assume without loss of generality $y_{a,b} = 1$, and observe that the no-cross constraint (10) forbids the selection of any arc (k,h) such that $(\{a,b\},\{k,h\}) \in \mathcal{C}$, hence $y_{k,h} = 0$ for all $(k,h) \in \mathcal{Q} \setminus \{(a,b),(b,a)\}$. This proves that y satisfies all inequalities (15), as claimed. □

Note that the improved no-cross constraints (15) can replace (10) in model (1)-(13), thus reducing their number from $O(|V|^4)$ to $O(|V|^3)$. However, although typically better, constraints (15) do not dominate (10) in the sense that a *fractional* solution y^* can violate a certain inequality (10) but no improved inequality (15). To see this, it is enough to consider a fractional solution y^* with $y^*_{a,b} = y^*_{b,a} = y^*_{h,k} = y^*_{k,h} = 1/3$, where $(\{a,b\},\{h,k\}) \in \mathcal{C}$ so (10) is violated, whereas $Q(a,b,k) = \{(a,b),(b,a),(k,h)\}$, and thus (15) is not violated.

2.4. Modeling obstacles and curvy cable connections

In practical applications, some obstacles can be present in the site, meaning that some areas of the site cannot be crossed by cables. As an illustration, consider the real-world offshore wind park of Figure 3 where two big obstacles are present—they are represented by the two polygons in the figure. All the given 29 turbines must be connected to the offshore substation, represented by the square node 0 in the figure. The substation has a limit of $C = 8$ connected cables. In this example we considered three types of cables:

- type 1 can support up to 5 turbines and has a price of 135 Euro/m;
- type 2 can support up to 7 turbines and has a price of 250 Euro/m;
- type 3 can support up to 9 turbines and has a price of 370 Euro/m.

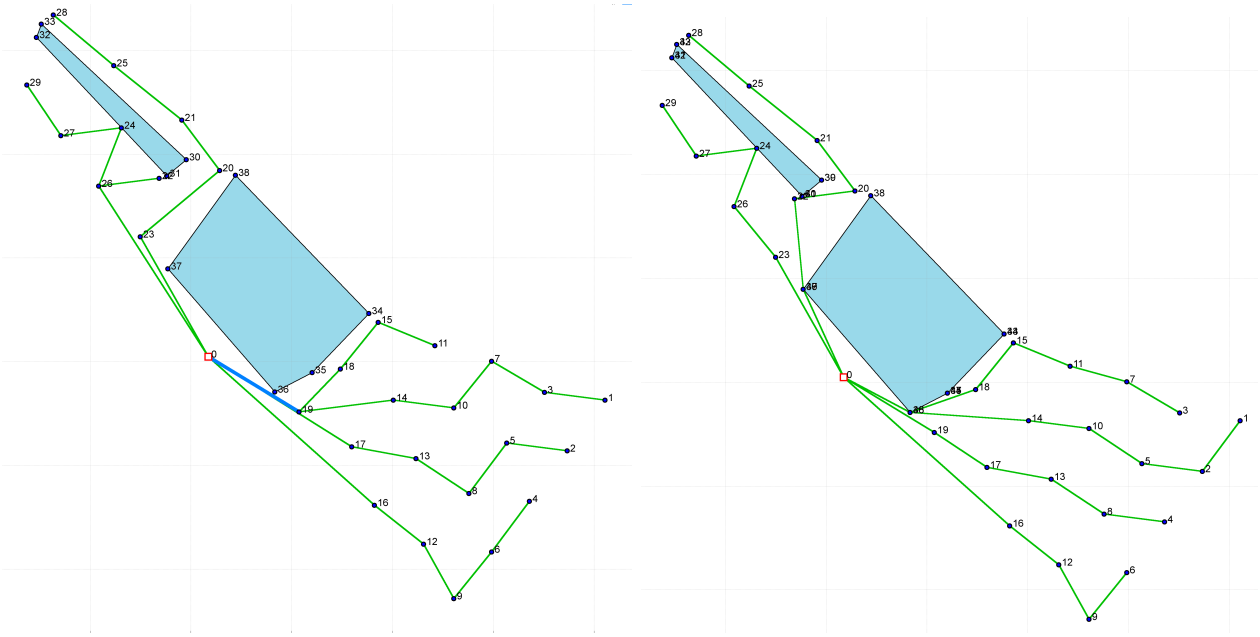
The prices above are publicly available and do not include installation costs. In Figure 3, type 1 cables are depicted in green, while type 2 cables are in bold blue (type 3 is not selected in any solution).

To define each obstacle area, viewed as a closed polygon, in our model we introduce artificial Steiner nodes (with no production) at the vertices of polygon and connect them through a circuit made by a artificial zero-cost cable that is forced in the solution by fixing the corresponding binary variable to 1. In this way, the final cost of the layout is not affected by the artificial cables but, due to the no-cross constraints, the actual cables cannot cross the obstacles. Referring to Figure 3 (left subfigure), only 9 Steiner nodes were introduced, namely:

- nodes 30, 31, 32, and 33, used to define the obstacle in the top-left corner.
- nodes 34, 35, 36, 37, and 38, used to define the big central obstacle;

Figure 3 (left) shows the optimized cable routing when considering only the 9 Steiner points above. Looking at the final layout, it is clear that the cable routing is rather expensive due to the

Figure 3 A real-world offshore wind farm with obstacles modeled through Steiner nodes. The layout on the right uses additional Steiner nodes to deal with curvy cable connections, and allows for a 7% cost saving.



non-flexibility of the cables (note, indeed, that our model can only deal with straight-line cable connections). For example, turbine 18 cannot be connected directly to the substation, because the

straight edge $[18, 0]$ would cross the obstacle. In this particular case, this has a significant impact on the final cost, as 18 is connected to 19 and therefore arc $(19, 0)$ needs a more expensive cable, since it has nine turbines connected; see the bold blue line in Figure 3 (left).

In real world, however, cables can be curvy so this connection would be possible. To simulate the curvy form of some cable connections, we can again use Steiner nodes. Our idea is to put some of them around the obstacles, so that the final solution can possibly use them to circumvent the obstacles. In particular, in the above example, one can put a new Steiner node over each of the artificial nodes delimiting the obstacles, plus two extra Steiner nodes overlying the two obstacle-points closer to the substation (which are in the most critical area). In this way a curvy cable can connect 14 to the substation and another can connect 18 to the substation. These connections support less flow (from five turbines each) and therefore can use a cheaper cable. Note, that two overlapping cables are not considered as a cable crossing since they can be laid in parallel. More specifically, in Figure 3 (right) the additional Steiner points used to add flexibility are:

- nodes 39, 40, 41, and 42 overlying respectively nodes 30, 31, 32, and 33;
- nodes 43, 44, 45, and 46 overlying respectively nodes 34, 35, 36, and 37;
- nodes 47, 48, and 49 overlying respectively nodes 35, 36 and 37.

Figure 3 (right) shows the new optimized cable routing, which uses only the less-expensive cable type 1 and allows for a cost saving of about 7% with respect to the previous one.

2.5. Modeling cable losses

In this section we extend the previous model to take cable losses into account. We first explain the physics behind the problem and next we introduce an alternative way to handle power losses through a simple preprocessing to be applied to the input data of our MILP model.

Physically, power losses are proportional to the square of the current. If we indicate with $g^t_{i,j} \geq 0$ the current actually passing through the cable of type t on arc (i, j), the total power loss in all the cable layout can be computed as

$$\sum_{(i,j)\in A} \sum_{t\in T} Q^t_{i,j} \, (g^t_{i,j})^2 \tag{16}$$

where $Q^t_{i,j}$ is a positive constant depending on the cable type and length that we define as

$$Q^t_{i,j} = R^t \cdot dist(i, j) \tag{17}$$

and R^t is the electrical resistance of cable type t, in Ω/m. If we want to estimate the value of these losses (in order to compare them with the layout price) we need to multiply the result by K_{euro}. K_{euro} is the cost for each MW of production.

Of course the current passing through each cable depends on the production of the connected turbines and is limited by the capacity of the used cable. Therefore, the new variables $g^t_{i,j}$ need to be linked to the remaining variables in the model through appropriate constraints.

As the electrical currents depend on the wind scenarios, one could think of a 2-stage Stochastic Programming model where x and y act as first stage variables, and we have a copy of variables $g^t_{i,j}$ for each wind scenario. This would lead to a really huge non-linear model, that would be very difficult to solve even for small instances.

Having understood the physic behind cable losses, we prefer to stick to a much simpler (and practically very effective) approach to deal with cable losses implicitly according to the following idea. We consider the MILP model without cable losses on a modified instance where each cable type is replaced by a series of *sub-cables* with discretized capacity and modified cable cost taking both installation costs and revenue losses into account.

As an illustration, consider a common situation where all turbines in the wind farm are identical, hence the maximum power production P_h of each turbine can be assumed to be 1. This means that we can express cable capacities as the maximum number of turbines supported by each cable type. Now consider a certain cable type t that can support up to k_t turbines. We replace it by k_t *sub-cable* types of capacity $f = 1, \ldots, k_t$ whose actual unit cost is computed by adding the cable/installation unit cost (u_t) and the unit power-loss cost (say $loss_{t,f}$) computed by considering the current produced by exactly f turbines. As unit costs increase with f, the optimal solution will always select the sub-cable type f supporting exactly the number of turbines connected, hence the approach models power losses in a correct way.

Note that the final sub-cable list would in principle contain $\sum_{t\in T} k_t$ different sub-cable types. However, for each sub-cable capacity $f = 1, \ldots, k_{MAX} = \max\{k_t : t \in T\}$ one only needs to keep the sub-cable with capacity f and minimum unit cost, i.e., only k_{MAX} "undominated" sub-cable types need to be considered in practice. This figure is not too large in practical cases, and can effectively be handled by our solution algorithms.

The above approach also allows us to consider multiple wind scenarios (and hence different current productions of the turbines) without the explicit need of second-stage variables, thus keeping the model size manageable. Indeed, one can just precompute the sub-cable unit costs by considering a weighted average of the loss cost under different wind scenarios. In our computational study, the wind energy experts computed the loss contribution to the unit cost for a 3-phase sub-cable of capacity f as

$$loss_{t,f} = 3 \sum_{s\in S} \pi_s (f I^s)^2 R^t K_{euro} \tag{18}$$

where S is the set of wind scenarios under consideration, π_s is the probability of scenario $s \in S$, and I^s is the current produced by a single turbine under wind scenario s assuming negligible wake effects, i.e., all turbines produce the same electric current.

Example: We will next show an example of how the power-loss prices are computed for a cable set named cb05 in Section 3.3 (these are realistic cables, though they do not refer to any specific cables on the market). We will consider the wind statistics from a real-world wind park in Denmark, namely Hors Rev 1 (named wf01 in Section 3.3).

Without cable losses, the cable cost would be taken directly from the cable information provided by the company, and would correspond to the sum of cable and installation costs—as reported in the last column of Table 1. When cable losses come into play, we need to modify the cable set and its prices, according to the strategy presented above. As cable type 1 supports up to 10 turbines, we need 10 sub-cables to deal with it, while 4 sub-cables are enough for cable type 2; see Table 2.

As to sub-cable prices, they need to consider also the power losses incurred under different scenarios. We used formula (18) with π_s and I^s taken from the actual wind statistics from the specific site. Parameter $K_{euro} = 0.68$ was computed by Vattenfall's experts by considering a cable lifetime of 25 years, a WACC of 8%, a warranted price of 0.10 Euro/KWh for 10 years and then a market price of 0.02 Euro/KWh, while resistance R_t is defined according to Table 1.

The prices considering cable cost, installation and losses for cb05 in the Horns Rev 1 case are shown in Table 2. Notice that, for each cable type, the sub-cable costs are monotonically increasing with the number of turbines supported, therefore in an optimal layout any sub-cable will support a number of turbines exactly equal to its maximum capacity.

For example, in row number 4 of Table 2 we consider the situation where cable type 1 is used to support the electrical current of 4 turbines. Its unit price is computed as 440 Euro/m for immediate costs, plus about 8.87 Euro/m for the estimated power losses related to the electrical currents produced by 4 turbines under the given wind scenarios. Note that the additional cost is non-linear, e.g., it is equal to about 33.54 Euro/m for 8 turbines. However, this non-linearity is handled in the input precomputation, without affecting the linearity of the underlying MILP model.

Table 1 Cable information for cb05.

cables	type	n. of 2MW turb. connected	resistance [Ohm/km]	cable price [Euro/m]	install. price [Euro/m]	total price [Euro/m]
cb05	1	10	0.13	180	260	440
	2	14	0.04	360	260	620

The comparison between Table 1 and 2 shows the impact of considering losses on cable prices. While from a installation perspective the cost for each cable type is fixed, now it varies depending

Table 2 Cable prices precomputed considering fixed costs and power losses for wf01 with cb05.

cable type	n. of 2MW turb. supported	price [Euro/m]
1	1	441.16831
	2	442.71000
	3	445.27949
	4	448.87676
	5	453.50184
	6	459.15471
	7	465.83537
	8	473.54382
	9	482.28007
	10	492.04412
2	11	639.77540
	12	643.41221
	13	647.36527
	14	651.63457

on how many turbines are connected. As we will see, this can have a large impact on the optimal cable routing.

2.6. No-cross constraint separation

As the number of no-cross constraints (15) on the complete graph $G = (V, A)$ can be very large for real-world instances, we decided not to include them in the model that is passed to the MILP solver. Instead, we generate them on the fly, during the MILP solver execution. To this end, we implemented a cut separation "callback" function that is automatically invoked by the solver to verify whether the current solution satisfies (15). This callback function is invoked both for (possibly fractional) solutions arising when solving the LP relaxation at any given branching-tree node, and for integer solutions generated by the internal MILP-solver heuristics.

Our separation function receives the (possibly fractional) solution y^* on input, and scans all node triples (a, b, k) with $a < b$ to check whether the corresponding improved no-cross constraint (14) is violated by y^*. Pairs (a, b) with $y^*_{a,b} = y^*_{b,a} = 0$ are skipped as they cannot lead to a violation constraint as y^* satisfies the out-degree inequalities (5)–(7).

Violated constraints, if any, are returned to the MILP solver in an appropriate format, so they can automatically be added to the current model. If no violated constraint (14) is found and y^* is fractional, we also apply a similar separation procedure to possibly generate violated constraints (14)(though this occurs in very rare cases).

3. Solution method

Since the MILP solver cannot solve large problems to optimality and often fails in even finding a feasible solution, we propose a matheuristic to find high-quality solutions in reasonable time.

3.1. A relaxed model

To be able to find meaningful (through possibly infeasible) initial solutions in very short computing times, we relax our MILP model by allowing it to produce disconnected solutions. To this end, for each $h \in V_T$ we replace the corresponding equality in (5) by a less-or-equal inequality. This however would not affect the final solutions due to the presence of the flow equilibrium equations (3). So we also relax the latter by allowing for a current loss in some nodes. This is obtained by introducing a slack continuous variable $l_h \geq 0$ for each $h \in V$, that indicates the energy lost at node h, and by replacing (3) by

$$\sum_{i \in V: i \neq h} (f_{h,i} - f_{i,h}) + l_h = P_h, \ h \in V_T \cup V_S \tag{19}$$

$$0 \leq l_h \leq k_{MAX}, \ h \in V_T \cup V_S \tag{20}$$

where $k_{MAX} = \max\{k_t : t \in T\}$ is the maximum cable capacity. The objective function coefficient of the new variables l_h is set to a very large positive constant, say $M \gg 0$, thus ensuring that a connected solution would always have a lower cost compared with a disconnected one.

To avoid solutions where no arcs are built, we add to the model a constraint that requires to install a total capacity in the arcs entering the substations that is sufficient to support the total turbine production, namely

$$\sum_{t \in T} \sum_{(i,j) \in A: j \in V_0} k^t x^t_{i,j} \geq \sum_{h \in V_T} P_h. \tag{21}$$

Our computational experience showed that a black-box MILP solver applied to the relaxed model above is typically able to find, in a few seconds, a feasible (possibly disconnected) first solution and to quickly proceed in the tree enumeration to discover better and better ones.

3.2. Matheuristics

As its name suggests, a *matheuristic* (Boschetti et al. 2009, Hansen et al. 2009) is the hybridization of mathematical programming with metaheuristics. The hallmark of this approach is the possibility of designing sound heuristics on top of a black-box MILP solver, by just changing its input data in a way that favors finding a sequence of improved solutions. In our settings, the black-box MILP solver is our exact method described in the previous section, where no-crossing constraints are separated on the fly as described in the next Section. // 2.6.

In our setting, the matheuristic is used as a refining tool that receives a certain solution H (described by its associated vector, say (x^H, y^H)) and tries to improve it by solving a restricted MILP "tailored around H" by fixing $y_{a,b} = 1$ for a suitable-defined subset of the y variables with $y^H_{a,b} = 1$. Note that this variable-fixing scheme is very powerful in our context as every time a certain $y_{a,b} = 1$ is fixed on input, one can forbid all possible crossing arcs, in a preprocessing phase, by just setting $y_{i,j} = 0$ for $(i,j) \in \mathcal{Q}(a,b,k) \setminus \{(a,b)\}$ for all $k \in V \setminus \{a,b\}$. This is very important for the success of our heuristic, as the restricted problem becomes much easier to solve due to the large number of variables fixed to 0 or to 1, and because of the fact that many relevant no-cross constraints are implicitly imposed by preprocessing.

In our implementation, we iteratively apply our refining matheuristic to the current best solution H available. At each iteration, we temporarily fix to 1 some y variables according to a certain criterion (to be described later), and apply the preprocessing described above to temporarily fix some other y variables to zero. We then apply the MILP solver to the corresponding restricted problem, and we warm start the solver by providing the current solution (x^H, y^H). We abort the execution as soon as a better solution is found, or a short time limit of a few seconds is reached. Then all fixed variables are unfixed, and the overall approach is repeated until a certain overall time limit (or maximum number of trials) is reached.

The very first solution H passed to the matheuristic plays an important but unpredictable role in determining the quality of the final solution. As a matter of fact, starting with a very bad (disconnected) solution can sometimes lead to very good final solutions. This behaviour suggests a *multi-start* strategy where a number of different initial solutions are generated and then iteratively improved by using our overall matheuristic. In our implementation, the initial solutions are defined by taking the very first (typically highly disconnected) solution found by the MILP solver when applied to our relaxed model with a random objective function where all the y variables have a random cost. (Note that the cost of the slack variables l_h introduced in (19) remains unchanged, meaning that disconnected solutions are still penalized.) To enhance diversification even further, we also use different input values for the MILP-solver's random seed parameter.

We next describe four possible variable-fixing criteria.

Our first criterion, called RANDOM, follows a simple *random variable-fixing* scheme that fixes variables $y_{a,b}$ with $y^H_{a,b} = 1$ with a certain probability, e.g., 50%.

Our second criterion, DISTANCE, uses a problem-specific strategy to choose the arc-fixing probability. To be more specific, the arc-fixing probability is related with the distance to the substation, namely: the arcs closer to the substation(s) are fixed with a larger/smaller probability.

Our third criterion, SWEEP, is specific for wind farms with only one substation. In these cases, the optimal solutions tend to divide the wind farm into *radial sectors* emanating from the substation. To take advantage of this property, we tentatively partition the wind farm into sectors,

and we then iteratively reoptimize each sector by fixing all the arcs that involve two nodes not belonging to it.

To be more specific, within SWEEP all nodes are initially ordered according to their angle with the substation. This produces a cyclic node sequence where the last-ranked node is followed by the first-ranked one. A *seed* node is randomly selected and a *sector* is defined by picking the $SPAN$, say, nodes that follow it in the cyclic ordered sequence. (In our implementation, we set $SPAN$ as a certain percentage of the total number of turbines.) The y variables associated with the arcs connecting two nodes outside this sector are fixed to their value in y^*, while any other y variable is left unfixed and the subproblem is reoptimized though our black-box MILP solver. If an improved solution is found, we select the seed s for the new iteration to be close to the previous one (i.e., in the cyclic sequence we randomly pick a node in the interval [s-2, s+2]). If the solution is not improved, instead, the new seed is taken by moving of 5 (say) nodes forward in the sequence. The SWEEP heuristic ends when the seed s moved along a complete cycle without improving the current solution.

Our fourth criterion, STRINGS, is similar to SWEEP but at each iteration it defines the sector to be reoptimized as follows. For each arc (i,r) with $y^*_{i,r} = 1$ that enters the substation, say r, we define the node set S_i containing all the nodes that reach the substation r passing through node i. In other words, S_i contains all the predecessors of node i in the anti-arborescence corresponding to y^*. Node sets S_i's are then sorted according to the angle formed by segment $[i,r]$ with respect to a vertical line passing through r. At each iteration, we choose a seed set S_{seed} and define the current *sector* to be optimized by taking $S_{seed} \cup S_{seed+1} \cup \cdots \cup S_{seed+SPAN}$ (where $SPAN$ is a given parameter and subscripts are taken in a cyclic way).

3.3. Computational tuning

To test how our four matheuristic algorithms perform on realistic data, we created a dataset of 11 synthetic wind farms. All instances are difficult as they involve a large number of turbines (from 60 to 93) and in some cases one or more obstacles.

Our matheuristics have been implemented on top of the state-of-the art commercial MILP solver IBM ILOG CPLEX 12.6. We used 3 instances for the parameter tuning of each algorithm (training set), and the remaining 8 to compare the matheuristics (test set). Each test has been run four times with four different random seeds, thus producing different final solutions. All matheuristics have been run with a time limit of 15 minutes on an Intel Xeon CPU X5550 running at 2.67GHz.

For the sake of space, we do not report here the detailed results of our experiments. Instead, we next summarize our findings.

- Matheuristic RANDOM is quite effective in its first iterations as it quickly finds good feasible solutions, though its performance becomes less satisfactory when the current solution becomes almost optimal.
- Matheuristic DISTANCE does not provide significantly better results than RANDOM, both when the fixing probability is larger for the arcs closer to the substation(s) and vice-versa. So RANDOM should be preferred as it is simpler and less prone to overtuning.
- Matheuristics SWEEP and STRINGS have a comparable performance and each of them tends to outperform RANDOM when the current solution is close to optimality
- The best overall results are obtained by using the above matheuristics in a combined way, as described in Section 3.4.

3.4. The overall algorithm

Our final algorithm is a mixture of matheuristic and exact solvers. To be specific, we first apply the RANDOM matheuristic with 50% fixing probability for 30 iterations, so as to quickly obtain a good solution. Then we apply STRINGS with $SPAN = 3$, and finally SWEEP with $SPAN = 0.3 \cdot |V_T|$. This matheuristic sequence is applied 5 times, in a multi-start vein, each time restarting from scratch from a different random initial solution. At each restart, the initial solution is defined as the very first (typically disconnected) solution found by the exact MILP solver (CPLEX 12.6) when applied to the relaxed model of Subsection 3.1 with a random objective function.

After the 5th restart, the exact MILP solver is applied to the original model without any heuristic variable fixing, using the best-available solution to warm-start the solver. In other words, our overall approach is intended to reach a proven optimal solution (if time limit permits). To this end, we first obtain very good heuristic solutions, in the matheuristic phase, and then we switch to the exact method to obtain a valid lower bound—and eventually to converge to proven optimality.

4. Computational analysis

This section is aimed at testing the performance of the MILP-based approach described in Subsection 3.4, using real-world data.

4.1. Test instances

For benchmark purposes we collected the data of five different real wind farms in operation in United Kingdom and Denmark. This data is available, on request, from the first author.

Table 3 summarizes the relevant information of the different sites for these five instances, named wf01, wf02, wf03, wf04 and wf05 in what follows. Figure 4 shows the different wind farms locations.

Figure 4 The real-world wind farms used in our tests.

We estimated the maximum number of connections to the substation, namely the input parameter C, by looking at the existing cable layout publicly available at (Kis-orca 2015) and (4cOffshore 2015). The corresponding constraint also determines the cable types to be used, which are reported in the last column of Table 3.

Our first wind farm (wf01) refers to Horns Rev 1, one of the oldest large-scale wind parks in the world. It was built in 2002 in the North Sea, about 15 km from the Danish shore, and produces around 160 MW. Our second wind farm (wf02) is Ketish Flats. Ketish Flats is located close to Kent in South East England, and can produce up to 140MW. It is a near-shore wind farm, so it is connected to the onshore electrical grid without any offshore substation. Nevertheless, only one export cable is connected to the shore, therefore we took the starting point of the export cable to act as a substation. This is handled by setting $C = \infty$ as there is no physical substation limitation in this case.

Our third wind farm (wf03) is Ormonde located in UK as well, but in the Irish Sea. It has a total capacity of 150 MW. Close to Kentish Flats (wf02) there is also Thanet (wf05), a bigger wind park with capacity 300 MW. When it was opened, in 2010, Thanet was the biggest offshore wind farm in the world. Finally wf04 refers to the DanTysk offshore wind farm, located west of the island of Sylt and directly on the German-Danish border. With a total of 80 wind turbines (288MW), DanTysk can provide up to 400,000 homes with green energy. All the considered sites are owned by Vattenfall.

The turbines in each wind park are of the same type, so we can assume $P_h = 1$ for all turbine nodes and we can express the cable capacities as the maximum number of turbines that it can support. Table 4 reports the capacity for the different cable types. We considered 5 different sets of real cables, named cb01, cb02, cb03, cb04, and cb05. As we know their capacity, resistance and

price, we could precompute their unit costs both with and without power losses, following the strategy we proposed in Subsection 2.5.

We already observed that power losses depend on the current flowing in the cables, that in turn depends on the average wind conditions within the site. We computed the cable-loss prices as a combination of real measured data and estimations based on Weibull distributions.

Table 3 Basic information on the real-world wind farms we used for tests.

name	site	turbine type	n. of turbines	C	allowed cables
wf01	Horns Rev 1	Vestas 80-2MW	80	10	cb01-cb02-cb05
wf02	Kentish Flats	Vestas 90-3MW	30	∞	cb01-cb02-cb03-cb04-cb05
wf03	Ormonde	Senvion 5MW	30	4	cb03-cb04
wf04	DanTysk	Siemens 3.6MW	80	10	cb01-cb03-cb04-cb05
wf05	Thanet	Vestas 90-3MW	100	10	cb04-cb05

Table 4 Basic information on the real-world cables we used for tests.

		n. of turbines supported			
cables	type	2MW	3MW	5MW	3.6MW
cb01	1	7	5	3	4
	2	11	8	4	6
	3	13	9	6	8
cb02	1	7	5	3	5
	2	12	8	5	7
cb03	1	12	8	5	7
	2	23	16	5	14
cb04	1	9	7	4	6
	2	21	15	9	13
cb05	1	10	7	4	6
	2	14	10	6	8

Each combination of site (i.e. wind farm) and feasible cable set represents an instance in our testbed, resulting in a total of 29 instances. Table 5 reports the main characteristics of our testbed. Each instance is identified by a number from 1 to 29 (first column in the table), and corresponds to an existing wind farm layout (second column) and to a set of possible cables (third column). Some cable sets refer to immediate costs of cables (i.e., CAPital EXpenditure, indicated as "capex" in the table) while others to costs including losses.

Table 5 Our testbed.

number	wind farm	cable set
01	wf01	wf01_cb01_capex
02		wf01_cb01
03		wf01_cb02_capex
04		wf01_cb02
05		wf01_cb05_capex
06		wf01_cb05
07	wf02	wf02_cb01_capex
08		wf02_cb01
09		wf02_cb02_capex
10		wf02_cb02
11		wf02_cb03
12		wf02_cb04_capex
13		wf02_cb04
14		wf02_cb05_capex
15		wf02_cb05
16	wf03	wf03_cb03_capex
17		wf03_cb03
18		wf03_cb04_capex
19		wf03_cb04
20	wf04	wf04_cb01_capex
21		wf04_cb01
22		wf04_cb03
23		wf04_cb04
24		wf04_cb05_capex
25		wf04_cb05
26	wf05	wf05_cb04_capex
27		wf05_cb04
28		wf05_cb05_capex
29		wf05_cb05

4.2. Tests

We first test the MILP model we developed in Section 2. To understand the capability of the model alone, we solve our real-world instances with a 10 hours time limit (Intel Xeon CPU X5550 running at 2.67GHz, CPLEX 12.6). The results are shown in Table 6. The first column indicates the instance considered (refer to Table 5 for details), the second the best solution found within the time limit, the third one reports the quality of the solution as computed by the solver and the last column reports the time used to find the proven optimal solution (36000.00 if the time limit is exceeded without finding it). It is seen that the MILP model alone works well for smaller instances: for instances 07 to 19 the optimal solution is found in a few seconds. However, these instances refer to the smallest wind farms in our data set (Ormonde and Kantish Flats). When considering bigger wind parks (as DanTysk and Thanet) the MILP solver alone is not able to find

even a feasible solution within 10 hours. Modern wind parks are often of the size of Thanet and DanTysk, so the algorithm should be designed to solve such instances.

Table 6 Solution quality obtained using our MILP model alone with a time limit of 36000 secs (– means infeasible solution).

Inst.	bestsol	% error LB	time
01	19437282.55	0.079	36000.00
02	–	–	36000.00
03	22611988.67	0.007	36000.00
04	24446239.21	0.470	36000.00
05	23482483.25	0.327	36000.00
06	–	–	36000.00
07	8555171.40	0.000	2.78
08	8806838.99	0.000	5.03
09	10056670.31	0.000	1.67
10	10303320.51	0.000	5.34
11	9200184.65	0.000	20.75
12	8604208.93	0.000	0.97
13	8933494.59	0.000	5.67
14	10173931.59	0.000	1.12
15	10348430.63	0.000	12.34
16	8054844.90	0.000	25.11
17	8560008.68	0.000	145.91
18	8357195.91	0.000	117.08
19	9178499.88	0.000	10029.88
20	–	–	36000.00
21	–	–	36000.00
22	–	–	36000.00
23	44421681.46	2.490	36000.00
24	–	–	36000.00
25	–	–	36000.00
26	22336016.56	3.330	36000.00
27	–	–	36000.00
28	–	–	36000.00
29	–	–	36000.00

Therefore, we now evaluate the capability of the matheuristic algorithm we presented in Subsection 3.4 to solve these real-world cable routing instances. To this end, we performed different runs with a time limit of 60, 300, 600, 1800, 3600 (1h), 36000 (10h), and 86400 secs (24h) on our computer.

Table 7 reports the quality of the solution found compared with the best solution known. For any instance, LB denotes the best-known lower bound computed by the exact solver after 24 hours of computing time. The table reports the instance number, the best-known feasible solution (best) and the associated percentage error with respect to LB (%error wrt best/LB), the computing time needed to get the best solution (time, in CPU secs), and then the percentage error with respect to LB and best at the various time limits. Solutions with a relative error of at most 0.01% satisfy the default optimality tolerance of our solver (10^{-4}), so they are considered optimal. Entries "–" refer to an infeasible (i.e., disconnected) solution found by the matheuristic in its early stage.

Table 7 Solution quality obtained using our matheuristic with different time limits. (– means infeasible solution).

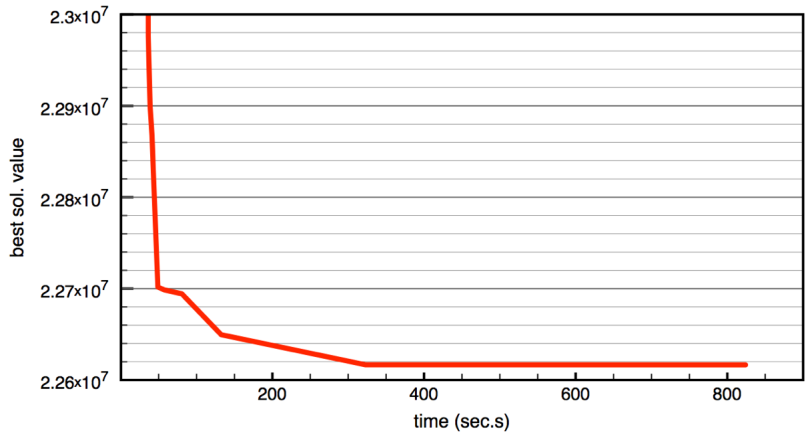
				60 s.		300 s.		600 s.		1800 s.		3600 s.		10 h		24 h	
		% error	time	%error wrt		%error wrt		%error wrt		%error wrt		%error wrt		%error wrt		%error wrt	
Inst.	best sol.	best/LB	to best	LB	best	LB	best	LB	best	LB	best	LB	best	LB	best	LB	best
01	19436700.18	0.01	14748.8	1.75	1.74	0.08	0.07	0.08	0.07	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00
02	21403410.11	0.09	86400.5	–	–	0.29	0.19	0.09	0.00	0.09	0.00	0.09	0.00	0.09	0.00	0.09	0.00
03	22611988.67	0.01	4621.3	0.39	0.38	0.18	0.17	0.03	0.02	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00
04	24445688.02	0.25	86400.9	–	–	0.45	0.21	0.27	0.02	0.25	0.00	0.25	0.00	0.25	0.00	0.25	0.00
05	23482483.25	0.17	86400.3	2.22	2.04	0.64	0.46	0.58	0.41	0.58	0.41	0.24	0.07	0.24	0.07	0.17	0.00
06	24768927.72	0.97	86401.3	2.85	1.86	1.01	0.04	0.97	0.00	0.97	0.00	0.97	0.00	0.97	0.00	0.97	0.00
07	8555171.40	0.00	43.0	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00
08	8806838.99	0.00	58.4	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00
09	10056670.31	0.00	28.3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	10303320.51	0.00	59.4	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00
11	9200184.65	0.00	174.4	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00
12	8604208.93	0.00	23.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
13	8933494.59	0.00	73.2	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00
14	10173931.59	0.00	24.4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
15	10348430.63	0.00	88.0	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00
16	8054844.90	0.00	247.8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
17	8560008.68	0.00	1278.3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
18	8357195.91	0.00	354.4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
19	9178499.88	0.01	1718.1	0.93	0.92	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00	0.01	0.00
20	38977593.84	5.10	86400.4	–	–	6.62	1.45	5.57	0.45	5.10	0.00	5.10	0.00	5.10	0.00	5.10	0.00
21	44857986.73	4.53	86401.3	–	–	4.88	0.34	4.53	0.00	4.53	0.00	4.53	0.00	4.53	0.00	4.53	0.00
22	40949573.29	2.11	86400.1	2.71	0.59	2.17	0.06	2.11	0.00	2.11	0.00	2.11	0.00	2.11	0.00	2.11	0.00
23	44421681.46	2.54	86400.1	4.89	2.29	2.54	0.00	2.54	0.00	2.54	0.00	2.54	0.00	2.54	0.00	2.54	0.00
24	50379247.34	7.94	86400.5	12.12	3.88	8.15	0.20	7.94	0.00	7.94	0.00	7.94	0.00	7.94	0.00	7.94	0.00
25	52331587.72	4.64	86400.9	–	–	9.13	4.29	7.43	2.67	6.59	1.87	4.71	0.06	4.64	0.00	4.64	0.00
26	22337935.84	3.49	86400.6	3.98	0.48	3.49	0.00	3.49	0.00	3.49	0.00	3.49	0.00	3.49	0.00	3.49	0.00
27	23362025.61	3.42	86400.2	5.52	2.03	3.85	0.42	3.45	0.03	3.42	0.00	3.42	0.00	3.42	0.00	3.42	0.00
28	26637602.25	2.57	86400.5	–	–	–	–	6.20	3.54	3.32	0.73	3.32	0.73	3.32	0.73	2.57	0.00
29	27295289.87	3.87	86401.7	–	–	–	–	–	–	4.84	0.94	3.87	0.00	4.09	0.21	4.09	0.22

Comparing Table 7 with the performance of the MILP model alone (Table 6), the impact of using a matheuristic framework is clear. Using this approach, indeed, we are able to provide feasible high quality solutions for all the instances, even the more difficult ones that were unsolvable earlier. According to Table 7, our method is able to solve 15 out of 29 instances to proven optimality (within the 0.01% tolerance), in many cases within few minutes. After just 1 minute, the method provides good solutions for all but 7 instances (all but 2 after 5 minutes, and all but 1 after 10 minutes). Very good (often provably optimal) solutions are available after 30 minutes for all the 29 instances in our testbed. After 1 hour the current solution value is, on average, just 0.03% from the best-known value, and 1.47% from the lower bound.

Figure 5 plots the best solution value over time for a typical run (instance n. 3). A feasible solution having objective 22,698,576.88 is found after 50 secs, within the initial matheuristic phase, while the best solution having objective 22,611,988.67 (at most 0.01% from optimum) is found after about 800 secs.

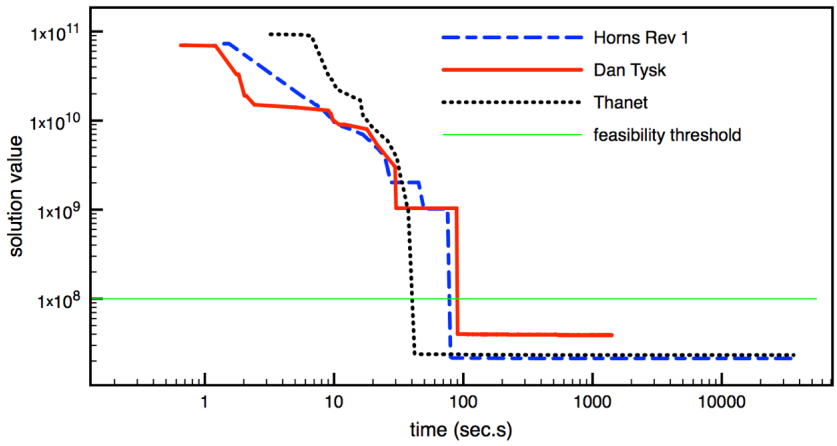
Note that our matheuristic framework is able to find feasible solutions very quickly also for the most difficult instances. In this matter, Figure 6 shows the evolution of the solution value for three difficult instances (2, 20 and 27) referring to three different real-world big parks (Horns Rev 1,

Figure 5 Heuristic solution value vs computing time for instance n. 3.



Dan Tysk and Thanet). These instances were unsolvable using the MILP solver alone, but using our matheuristic approach a feasible solution is found in about 100 seconds.

Figure 6 Heuristic solution value vs computing time for instances n. 2, 20 and 27. The green line indicates the feasibility threshold: solutions under the green line are feasible (connected) solutions. These three instances, unsolvable with the MILP solver alone, now reach a feasible solution in less than 100 seconds.

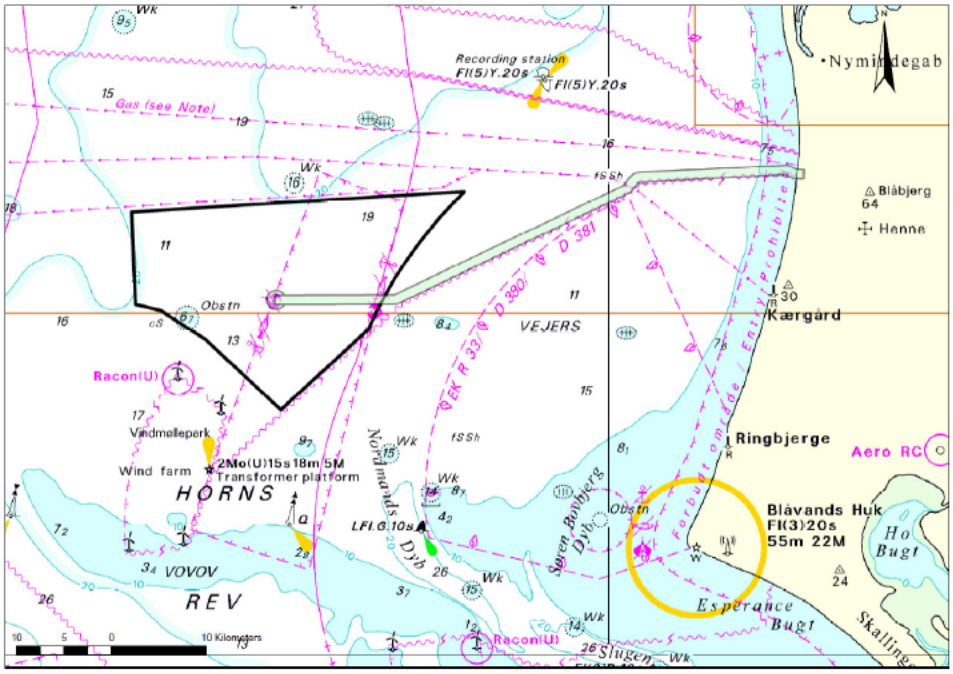


5. Discussion of a real case

In the previous sections we discussed the methodology we used to optimize offshore cable routing and its effectiveness for real-world cases. In what follows, we would like to give an example of the impact of considering losses on the cable layout solution. This is a major result from a commercial perspective since, having a cable route optimized considering power losses, can translate into huge

savings. Here, we refer to a new wind farm, Horns Rev 3 (HR3) that is still under construction (Energinet.dk 2013). The planned Horns Rev 3 is an area of approximately $160km^2$ in the eastern North Sea, 10-20km north of Horns Rev.

Figure 7 Location of Horns Rev 3 (Energinet.dk 2013)



Energinet.dk has designed the substation for the future offshore wind farm as well as the connections with the coast. Therefore, the position of the substation and its characteristics are fixed. We considered a preliminary layout provided by Vattenfall with 50 8MW turbines, where turbine positions are fixed. The project team already decided that the following types of cable must be used for our tests:

- type 1 can have up to 3 turbines connected, a price of 180 Euro/m, a resistance of 0.13 Ohm/km, and insulation losses of 111 W/km.
- type 2 can have up to 4 turbines connected, a price of 360 Euro/m, a resistance of 0.04 Ohm/km, and insulation losses of 109 W/km.

The substation can be connected to only 12 cables, therefore we need at least 2 cables of 5 turbines each. For this reason we were allowed to overload two of the cables of type 2 (in this case it would carry 5 turbines each). Note that these overloaded cables need particular systems to monitor their temperature, which are expensive, therefore we had to limit their use (they can be used only two times).

We considered an installation price fixed at 260 Euro/m for both types of cable. The problem has been studied considering 13 real-world wind scenarios from HR3 with 8MW turbines. From this data, we compute the cost of cables considering losses according to (18). Considering cable prices, installation, cable losses in 25 years and insulation losses, the equivalent prices for each type of cables are shown in Table 8.

Table 8 Costs for cable type 1 (considering power losses over 25 years) supporting 1, 2 or 3 turbines respectively. Costs for cable type 2 (considering losses over 25 years) supporting 4 or 5 turbines respectively

	1wtg	2wtg	3wtg	4wtg	5wtg
cable 1	470.9	561.3	711.87		
cable 2				786.5	879.7

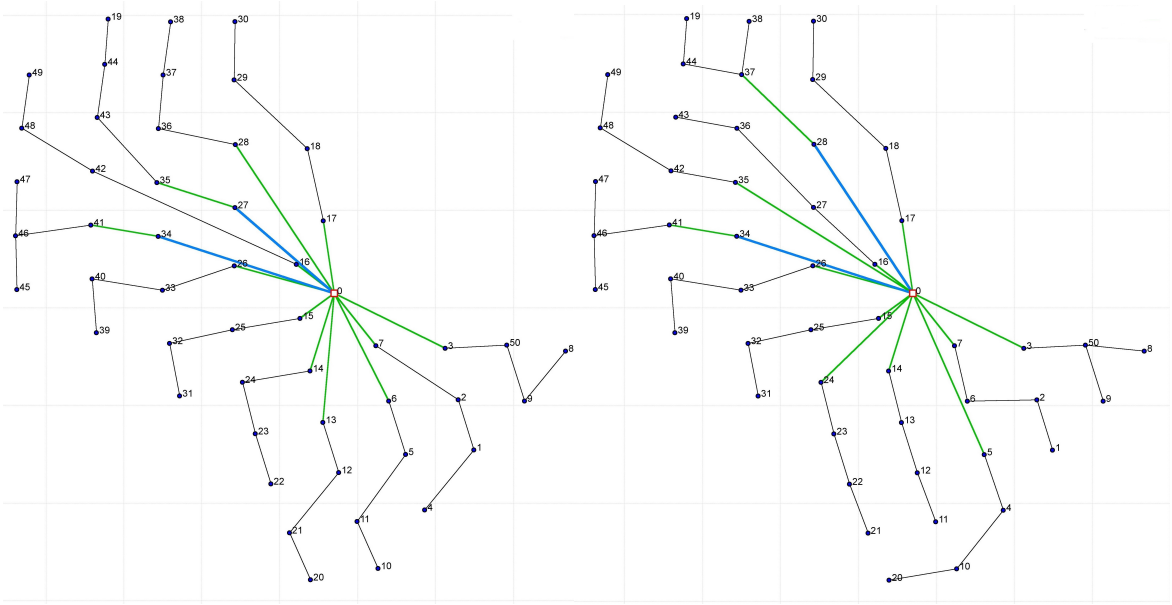
We ran our matheuristic with the original cable prices and installation costs (without losses) according to model (1) – (13). The solution found is therefore an optimized cable route from an immediate investment perspective, the so called CAPital EXpenditure (CAPEX). When we considered the new prices of Tables 8, we obtained a different layout: this is the optimal layout from a long term perspective, and is the one that would allow a bigger revenue in 25 years. The two layouts are compared in Figure 8.

In this case the long-term-optimized layout costs around 370KEuro more at construction time but after 25 years this amount is paid back and another 891KEuro (net present value) are saved.

One could also analyze the structure of this long-term-optimized solution compared with the CAPEX optimized one. First, in the layout optimized considering losses (second plot in Figure 8) there are more nodes with in-degree larger than 1, i.e., with more than one cable entering a turbine. In the layout we have 3 such nodes versus only one of the CAPEX optimized layout (first plot in Figure 8). To understand why, let us e.g. observe the connection between turbines 50, 8 and 9. Arc (8,9) is slightly shorter than arc (8,50), therefore it is more convenient from a CAPEX perspective. However, when considering the losses, we notice that in cable (9,50) of the first layout in Figure 8 the current from 2 turbines is flowing, therefore there is a larger loss compared to the connection (9,50) of the second layout where only one turbine is connected. All in all, this structure is more convenient from a long-term perspective because it maximizes the use of cable that carries less energy (from only one turbine).

We can finally compare the use of the different types of cable in the different solutions (Table 9): in the layout considering only CAPEX, 63% of the total cable length is for type 1 cable (the

Figure 8 Optimized layouts for HR3: On the left is the layout optimized by considering only CAPEX costs, while on the right is the optimal layout considering cable losses over 25 years. Black cables are of cable type 1, green are of cable type 2, and blue are of cable type 2 carrying 5 turbines.



(a) CAPEX optimized layout

(b) Long-term optimized layout

Table 9 Use of different types of cable in the two solutions. The first table refers to the CAPEX optimized solution, where 63% of the total cable length is in smaller (less expensive) cables. The second table refers to the optimized layout considering losses: the use of thicker cables (less sensible to losses) increased.

CAPEX optimized layout		
type	length [m]	% of the total
1	57149	63
2	33387	37

Long-term optimized layout		
type	length [m]	% of the total
1	52715	59
2	37130	41

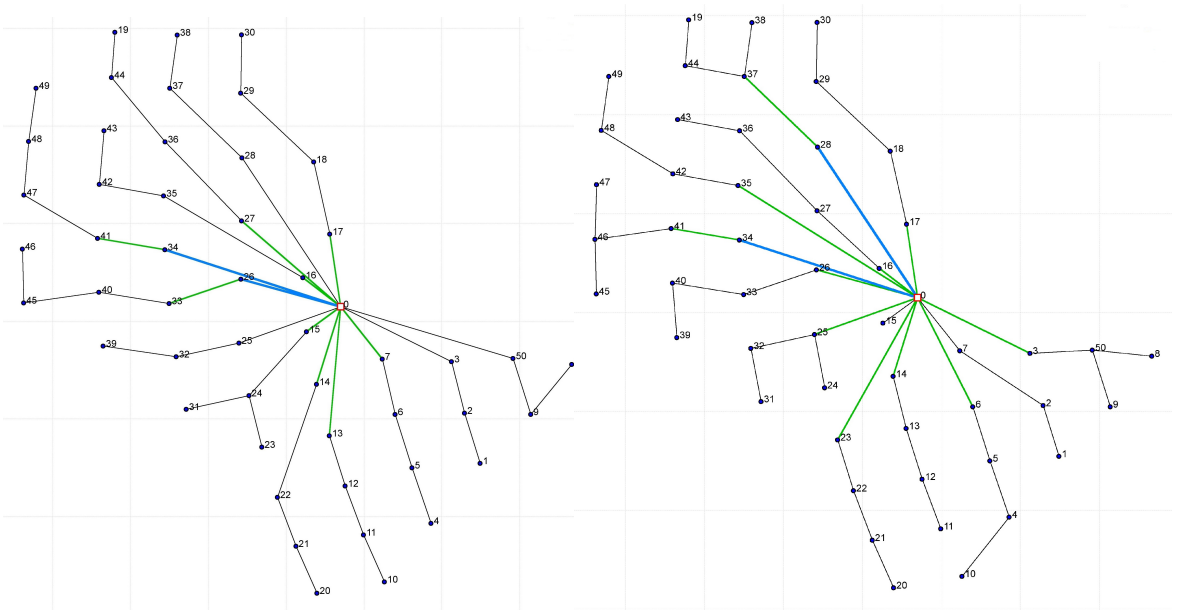
cheapest), while in the layout considering losses this value decreases to 59%. This reflects the fact that the smallest (i.e. the cheapest) cable has the highest losses.

This real-world test case was particularly limited by the number of connections to the substation, therefore we re-ran the cable routing optimization for the same site but imposing a maximum of 18 connections (instead of 12). Of course, to physically add connections can significantly increase the substation price, but here we just want to test the optimizer behaviour, so we suppose to have a substation already built with the extra connections. In this test the optimization is more free to play with the different cable types.

The results of the optimization considering CAPEX prices only, as well as those considering losses over 25 years (same prices as in Table 8) are shown in Figure 9.

The new layouts confirm the tendencies already discussed: the layout optimized considering losses has more nodes with higher degree than the other, and is using more expensive cables. Table 10

Figure 9 Optimized layouts for HR3 with maximum 18 connections to the substation: plot on the left is the optimal layout considering only CAPEX costs while on the right is the layout optimized considering cable losses over 25 years. Black cables represent cable type 1, green are cable type 2 and blue are cable type 2 overloaded.



(a) CAPEX optimized layout (b) Long-term optimized layout

shows the use of different types of cable in the two new solutions: in this case there is a difference of 15% in the use of the different kinds of cable.

Table 10 Use of different types of cable in the two solutions. The first table refers to the CAPEX optimized solution, where 74.5% of the total cable length is in smaller (less expensive) cables. The second table refers to the layout optimized by considering losses: the use of thicker cables (less sensible to losses) increased.

CAPEX optimized layout		
type	length [m]	% of the total
1	68910	74.5
2	23584	25.5

Long-term optimized layout		
type	length [m]	% of the total
1	53447	59
2	36693	41

In the 18 connections case the long-term layout costs around 1 324 KEuro more at construction time but in 25 years this amount is paid back and another 1 156KEuro (net present value) are saved. All in all, these results show the importance of considering losses when designing the inter-array cable routing.

6. Conclusions

In this paper we have introduced a new MILP model for optimal cable routing in offshore wind farms. A main novelty of our model is its capability of taking both installation costs and power

losses into full account. We have also developed a new matheuristic framework in order to have a practical optimization tool to be used on difficult real cases. Thanks to the collaboration between Vattenfall and DTU, we have been able to describe the problem as it appears in real applications, and to validate our results on real-world instances that have been made publicly available for benchmarking. Using a sound matheuristic framework, for most of our instances we have been able to produce extremely good solutions in about 15 minutes of computing time on a standard PC.

Finally, we have fully analyzed a real wind farm whose data was provided by Vattenfall, namely Horns Rev 3, to quantify the commercial impact of considering power losses in the cable routing design. We have shown that savings in the order of millions Euros can be achieved in a wind park lifetime. These kinds of highly-optimized layouts cannot be produced by a manual operator, due to the complexity of the corresponding design problem. In addition, evaluating the impact of losses on a long term perspective and understanding how the layout should be changed in order to reduce them, represents an extremely valuable analysis for a company, that could not have been carried out otherwise.

Acknowledgments

This work is supported by Innovation Fund Denmark. Thanks to Jesper Runge Kristoffersen, Iulian Vranceanu, Thomas Hjort and Kenneth Skoug from Vattenfall BA Wind who helped us in defining the cable routing constraints.

References

4cOffshore. 2015. website. `http://www.4coffshore.com/offshorewind/`. Accessed: 2015-01-30.

Archer, R., G. Nates, S. Donovan, H. Waterer. 2011. Wind turbine interference in a wind farm layout optimization – mixed integer linear programming model. *Wind Engineering* **35, no.2** 165–178.

Bauer, J., J. Lysgaard. 2015. The offshore wind farm array cable layout problem: a planar open vehicle routing problem. *Journal of the Operational Research Society* **66**(3) 360–368.

Berzan, C., K. Veeramachaneni, J. McDermott, U.M. O Reilly. 2011. Algorithms for cable network design on large-scale wind farms. *Tech. Rep. Tufts University* .

Boschetti, M.A., V. Maniezzo, M. Roffilli, A. Röhler Bolufé. 2009. Matheuristics: Optimization, Simulation and Control. *Hybrid Metaheuristics*. Springer Berlin Heidelberg, 171–177.

Cerveira, A., A. De Sousa, E. J. S. Pires, J. Baptista. 2016. Optimal cable design of wind farms: The infrastructure and losses cost minimization case. *IEEE Transactions on Power Systems* (31.6) 4319 – 4329.

Cordone, R., F. Maffioli. 2004. On the complexity of graph tree partition problems. *Discrete Applied Mathematics 134* 51–65.

Dutta, S. 2012. Data mining and graph theory focused solutions to smart grid challenges. Master's thesis, University of Illinois.

Dutta, S., T. J. Overbye. 2011a. A clustering based wind farm collector system cable layout design. *Power and Energy Conference at Illinois (PECI)*. 1–6.

Dutta, S., T. J. Overbye. 2011b. A clustering based wind farm collector system cable layout design. *Power and Energy Conference at Illinois (PECI), 2011 IEEE*. 1–6.

Energinet.dk. 2013. Technical project description for the large-scale offshore wind farm (400mw) at horns rev 3, document no. 13-93461-267 .

Fagerfjall, P. 2010. Optimizing wind farm layout - more bang for the buck using mixed integer linear programming. Master's thesis, Department of Mathematical Sciences, Chalmers University of Technology and Gothenburg University, Goteborg, Sweden.

Fischetti, Martina, Michele Monaci. 2015. Proximity search heuristics for wind farm optimal layout. *Journal of Heuristics* doi:10.1007/s10732-015-9283-4.

Garey, Michael R., David S. Johnson. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman.

González, Javier Serrano, Manuel Burgos Payán, Jesús Manuel Riquelme Santos, Francisco González-Longatt. 2014. A review and recent developments in the optimal wind-turbine micro-siting problem. *Renewable and Sustainable Energy Reviews* **30** 133–144.

González-Longatt, F. M., P. Wall. 2012. Optimal electric network design for a large offshore wind farm based on a modified genetic algorithm approach. *IEEE Systems Journal* (6.1) 164–172.

Hansen, P., V. Maniezzo, S. Voß. 2009. Special issue on mathematical contributions to metaheuristics editorial. *Journal of Heuristics* **15**(3) 197–199.

Hertz, A., O. Marcotte, A. Mdimagh, M. Carreau, F. Welt. 2012. Optimizing the design of a wind farm collection network. *INFOR* **50**(2) 95–104.

Kis-orca. 2015. website. `http://www.kis-orca.eu/downloads/`. Accessed: 2015-01-30.

Kusiak, Andrew, Zhe Song. 2010. Design of wind farm layout for maximum wind energy capture. *Renewable Energy* **35**(3) 685–694.

Li, D.D., Chao He, Yang Fu. 2008. Optimization of internal electric connection system of large offshore wind farm with hybrid genetic and immune algorithm. *Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT2008)*. 2476–2481.

Qi, Wie, Yong Liang, Zuo-Jun Max Shen. 2015. Joint planning of energy storage and transmission for wind energy generation. *Operations Research* **63**(6) 1280–1293.

Samorani, Michele. 2013. The wind farm layout optimization problem. *Handbook of Wind Power Systems*. Energy Systems, 21–38.

Zhao, M., Z. Chen, F. Blaabjerg. 2009. Optimisation of electrical system for offshore wind farms via genetic algorithm. *IET Renewable Power Generation 3* (3.2) 205–216.