



What matters for wind farm layout optimization? An assessment based on the effective layout distance

Matteo Baricchio¹, Jenna Iori², Pieter M.O. Gebraad³, and Jan-Willem van Wingerden¹

¹Faculty of Mechanical Engineering, Delft University of Technology, Delft, the Netherlands

²Faculty of Aerospace Engineering, Delft University of Technology, Delft, the Netherlands

³Youwind, Barcelona, Spain

Correspondence: Matteo Baricchio (m.baricchio@tudelft.nl)

Abstract. Wind farm layout optimization is one of the main phases of offshore wind farm design and involves a wide range of implementation choices, spanning aspects such as flow modeling, data selection, and objective definition. This study quantifies the impact of 18 different implementation choices across 9 categories on the results of wind farm layout optimization. The comparison is performed using both a conventional approach based on performance metrics and the concept of *effective layout distance*, a novel metric introduced to measure differences between sets of layouts based solely on turbine positions. The results show that the hyperparameter values of the optimization algorithm, the choice of wake model, the use of cost of valued energy as objective, and the adoption of a control co-design approach have a high impact on the optimal layouts and therefore require careful consideration during problem formulation. In contrast, choices such as rotor averaging models and the use of revenues based on varying electricity prices as objective have a negligible effect on the resulting layouts and can therefore be selected based on simplicity and computational efficiency.

1 Introduction

Wind farm layout optimization (WFLO) is a key component of the design phase of offshore wind farms, consisting in the placement of turbines within a defined area to maximize overall performance (Mosetti et al., 1994). This process typically aims to minimize wake losses, thereby increasing the annual energy production (AEP). The outcome of WFLO is the set of turbine coordinates defining the layout. Several implementation choices arise in this process, including the selection of flow models, input data, objective functions, and optimization algorithms. As a result, WFLO inherently requires a trade-off between model accuracy and computational efficiency (Réthoré et al., 2014).

In WFLO, the flow is commonly modeled using engineering wake models, which enable fast simulation of wake interactions between turbines while providing a sufficient level of accuracy (Zhang et al., 2026). These models capture the impact of wake deficit, wake superposition and rotor average speed calculation on the power production, and are implemented in widely used software packages such as FLORIS (NatLabRockies, 2026) and PyWake (Pedersen et al., 2023). In recent years, several wake deficit models have been proposed (Porté-Agel et al., 2020; Göçmen et al., 2016). These models differ in complexity, incorporating an increasing number of physical effects, such as the dependence on turbulence, and thus involve varying computational cost. However, no clear consensus exists on which model provides the highest accuracy (Jong et al., 2026). This is partly



25 because these models differ not only in formulation but also in tunable parameters, which can significantly affect the resulting wake characteristics (Guo et al., 2021). An example of these parameters is the wake expansion coefficient, present in most deficit models and tuned according to higher-fidelity simulations or experimental data. Consequently, the WFLO literature employs a wide range of wake deficit models, with no clear consensus on the appropriate trade-off between model accuracy and computational performance (Gao et al., 2020).

30 Few studies have compared the use of different wake deficit models in WFLO, showing that the robustness of the resulting layouts varies depending on the specific model and case study when cross-evaluated using alternative models (Yang and Najafi, 2021; Romero-Barrera et al., 2025). This implies that layouts optimized with a given model can become suboptimal when an alternative model is used for their evaluation. O'Neill et al. (2025) highlighted the uncertainty faced by designers when selecting a wake model and proposed a multi-objective approach to mitigate this issue by simultaneously optimizing AEP predictions
35 from multiple models, including those by Jensen (1983), Bastankhah and Porté-Agel (2014), and Nygaard (2022). Similarly, Baigarina et al. (2026) demonstrated the significant influence of the wake expansion coefficient on WFLO outcomes. Beyond wake deficit models, several additional modeling choices can influence WFLO results, including wind resource representation, wake superposition methods, rotor-averaging approaches, and flow discretization defined by wind direction and wind speed binning. These choices involve different trade-offs between accuracy and computational cost; however, the literature does not
40 provide clear criteria for selecting among them or establish the extent to which they can affect the resulting WFLO solutions.

Another aspect that differentiates WFLO studies in the literature is the selection of the performance metric used as the objective function. AEP is the most commonly used objective, as it is equivalent to revenues under traditional subsidy schemes, while requiring relatively few assumptions and, in some cases, directly providing gradient information (Azlan et al., 2021). Alternatively, revenue-based objectives can account for variations in electricity prices across wind conditions (Dirik et al.,
45 2025). Cost-related metrics, such as the levelized cost of energy (LCOE) and net present value (NPV), are also employed to capture the trade-off between energy production and cost of the asset (Zhang et al., 2026). However, the criteria used to select one objective over another are not clearly established in the literature, and the consequences of this choice for the resulting turbine layouts remain unexplored.

Recent research has introduced co-design concepts for the WFLO problem, in which turbine positions are optimized simul-
50 taneously with additional design or operational variables instead of sequentially. Examples include the joint optimization of cable layouts and turbine positions (Sanchez Perez Moreno, 2019), as well as the integration of wind farm control strategies into the layout optimization process (Stanley et al., 2023). To address the resulting increase in problem dimensionality, simplified models and approximations are often employed, such as heuristic solvers for cable design or geometric relations for turbine control setpoints (Souza de Alencar et al., 2025; Baricchio et al., 2024; Italiano et al., 2026). These approaches have
55 demonstrated performance improvements compared to traditional sequential optimization strategies for selected scenarios. For instance, in the context of control-layout co-design, co-design can increase AEP by up to 0.8% (Stanley et al., 2023); however, the magnitude of this gain appears to depend strongly on the specific scenario and modeling choices, with no improvement observed in some cases (Baricchio et al., 2024, 2026a).



Another key aspect of WFLO is the choice of optimization algorithm. The challenging nature of the WFLO problem requires optimization algorithms that can efficiently explore the design space and converge towards high-quality solutions. As the size of the wind farm increases, WFLO becomes a high-dimensional optimization problem, with the number of design variables equal to twice the number of turbines (Poole, 2026). Moreover, the problem is multimodal, with multiple local optima that can hinder algorithm convergence (Stanley and Ning, 2019). Consequently, several optimization methods tailored to WFLO have been proposed to improve the efficiency and reliability of the optimization process (Azlan et al., 2021). Thomas et al. (2023) compared several widely used algorithms and showed that multiple approaches can achieve good performance when properly configured. This highlights the importance not only of algorithm selection but also of hyperparameter tuning.

The literature demonstrates that numerous implementation choices can influence WFLO. Previous studies have examined the impact of specific categories, such as wake models or optimization algorithms, but it remains unclear which choices are most impactful. Thomson and Zaaier (2022) comprehensively assessed the impact of alternative WFLO implementations by comparing the AEP obtained with each implementation against that of a reference. Their analysis considered variations in wake deficit models, wake superposition methods, wind resource representations, and flow discretization. They found that most alternatives had a limited effect on AEP when evaluated against the reference, often within the variability observed across multiple optimization runs. The largest impacts were associated with optimization algorithms and wind resource representation, although the latter involved strong simplifications.

However, the impact of different implementation choices on the turbine positions remains unclear. Most existing studies assess implementation choices based on performance metrics, typically computed using the same low-fidelity models adopted within the WFLO. This raises the question of whether the observed trends persist when layouts are evaluated using higher-fidelity models. Therefore, to avoid such bias, a more effective approach would be to directly assess differences in turbine positions by visually comparing the resulting layouts. However, as the size of the wind farm and the number of layouts to be compared increase, this approach becomes impractical, as it requires the visual inspection of a large number of turbine positions distributed over an extended area. Moreover, the multimodal nature of the WFLO problem and the stochasticity of the optimization algorithms typically used to solve it further complicate the analysis, as repeated runs of the same optimization can produce substantially different layouts (Poole, 2026). As a result, comparisons often involve sets of layouts rather than individual solutions (Thomson and Zaaier, 2022). Focusing only on the best-performing layout may introduce bias due to the low-fidelity model used to select such a layout and may not capture the overall variability of the solution space. Comparing sets of layouts further complicates the visual assessment of turbine positions and the identification of the impact of different implementation choices, particularly due to the multimodal nature of the WFLO problem. In this context, small differences in performance metrics can correspond to substantial differences in layout configurations.

This study addresses this gap by comparing layouts obtained under different implementation choices using a metric termed *effective layout distance*, derived from the concept of layout field error introduced by Baricchio et al. (2026b).

The main contributions of this study are:

- Quantifying the impact of key implementation choices in WFLO on turbine positions.



- Comparing these impacts to identify which implementation choices can lead to substantially different layouts and therefore require careful consideration, and which have a negligible effect on the resulting layouts compared to simpler alternatives.

95

The remainder of the paper is structured as follows. Section 2 presents the proposed methodology, which defines a systematic approach to assess the impact of different implementation choices. The proposed approach considers two complementary comparison methods: first, a comparison based on performance metrics, which enables consistency with the state-of-the-art approach, and second, a comparison based on turbine positions, introducing the effective layout distance as a metric to quantify differences between layouts. Section 3 describes the case studies, specifying the implementation choices that are investigated and the selected wind farm characteristics. The results are presented in Sect. 4 and discussed in Sect. 5. Finally, conclusions and recommendations for future work are provided in Section 6.

100

2 Methodology

105

This section describes the methodology adopted in this study. Section 2.1 first presents the formulation of the WFLO problem considered in this work and identifies the stages at which different implementation choices arise. Section 2.2 then describes the traditional assessment based on performance metrics, while Sect. 2.3 introduces the effective layout distance metric and the proposed assessment based on turbine positions.

2.1 Wind farm layout optimization problem

110

The WFLO problem is formulated as the maximization of an objective function f , as defined in Eq. (1). The design variables are the turbine positions, represented by N_{wt} continuous Cartesian coordinates $(\mathbf{x}, \mathbf{y})$. These positions are constrained to lie within a predefined area Ω and to satisfy a minimum pairwise distance d_{min} .

$$\begin{aligned} & \max_{\mathbf{x}, \mathbf{y}} f(\mathbf{x}, \mathbf{y}) \\ & \text{subject to } (\mathbf{x}, \mathbf{y}) \in \Omega \\ & \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \geq d_{\text{min}} \quad \forall i, j \in [1, N_{\text{wt}}], i \neq j \end{aligned} \quad (1)$$

115

The objective function f is a performance metric that depends on the turbine positions $(\mathbf{x}, \mathbf{y})$, and can be affected by several implementation choices. A detailed list and description of the implementation choices considered in this study is provided in Sect. 3.1. These include modeling choices, such as wake model selection and wind resource discretization, which lead to different estimates of the same metric (e.g., AEP computed with different wake models). The choice of objective function, such as AEP or LCOE, or the formulation as a co-design problem, are also relevant implementation choices. Lastly, the choice of the optimization algorithm constitutes a separate implementation choice, not directly related to the definition of f .

120

In the remainder of this paper, the following notation is used. A layout $(\mathbf{x}, \mathbf{y})$ obtained using an implementation choice A is denoted by $\mathbf{X}_A$, while the set of layouts obtained from N_{runs} independent optimization runs is denoted by $\mathcal{L}_A$. Similarly, when an implementation choice B is used, the corresponding layout and set are denoted by $\mathbf{X}_B$ and $\mathcal{L}_B$, respectively.



Following the approach of Thomson and Zaaier (2022), a referent is selected among the implementation choices considered in this study to provide a common basis for their comparison and classification. It is denoted by REF and described in Sect. 3.1.

2.2 Layout comparison based on performance metrics

Two layouts, $\mathbf{X}_A$ and $\mathbf{X}_B$, can be compared based on their performance metrics. As proposed by Romero-Barrera et al. (2025), a cross-evaluation is performed in which both layouts are evaluated using f_A and f_B , with the subscript denoting the implementation choice embedded in the performance metric. We distinguish between *design choice* and *evaluation choice*, depending on whether the implementation choice is adopted during the layout optimization or the evaluation stage. This approach is extended to the sets $\mathcal{L}_A$ and $\mathcal{L}_B$, providing a statistical basis for their comparison. The remainder of this section describes the approach adopted in this study for comparing layout sets based on performance metrics, which consists of four sequential steps, and is also illustrated in Fig. 1:

1. *Wind farm layout optimization*: The WFLO problem is solved N_{runs} times for each implementation choice A and B , i.e. design choices, resulting in two sets of layouts, $\mathcal{L}_A$ and $\mathcal{L}_B$.
2. *Model evaluation*: Each layout in $\mathcal{L}_A$ and $\mathcal{L}_B$ is evaluated using both performance metrics f_A and f_B . This cross-evaluation results in $4 \times N_{\text{runs}}$ performance metric values. For clarity, the notation $F_{A|A}$, $F_{B|A}$, $F_{A|B}$, and $F_{B|B}$ is used to indicate the design and evaluation choices, respectively. Specifically, $F_{A|A}$ and $F_{B|A}$ correspond to the layouts obtained with design choice A , evaluated using f_A and f_B , respectively. Similarly, $F_{A|B}$ and $F_{B|B}$ correspond to the layouts obtained with the design choice B .
3. *Median calculation*: The median value is computed for each of the four combinations, condensing the N_{runs} performance values into a single representative value. The resulting median values are denoted a tilde and given by $\tilde{F}_{A|A}$, $\tilde{F}_{B|A}$, $\tilde{F}_{A|B}$, and $\tilde{F}_{B|B}$.
4. *Normalization*: To compare implementation choices A and B , the performance metrics are normalized using as a reference the case in which the implementation choice used during optimization and evaluation is identical, namely $\tilde{F}_{A|A}$ and $\tilde{F}_{B|B}$. The normalization is applied both to the individual performance values and to their corresponding median values. For example, the percentage difference between evaluating layouts obtained with B using implementation choice A and the corresponding matched case is calculated as

$$\Delta_{A|B-A} = 100 \cdot \frac{F_{A|B} - \tilde{F}_{A|A}}{\tilde{F}_{A|A}}. \quad (2)$$

Similarly, the percentage difference between the median values is computed as

$$\tilde{\Delta}_{A|B-A} = 100 \cdot \frac{\tilde{F}_{A|B} - \tilde{F}_{A|A}}{\tilde{F}_{A|A}}. \quad (3)$$

The same procedure is applied to the opposite comparison, resulting in $\Delta_{B|A-B}$ and $\tilde{\Delta}_{B|A-B}$. Consequently, N_{runs} values are obtained for both $\Delta_{A|B-A}$ and $\Delta_{B|A-B}$, together with a single value for both $\tilde{\Delta}_{A|B-A}$ and $\tilde{\Delta}_{B|A-B}$.

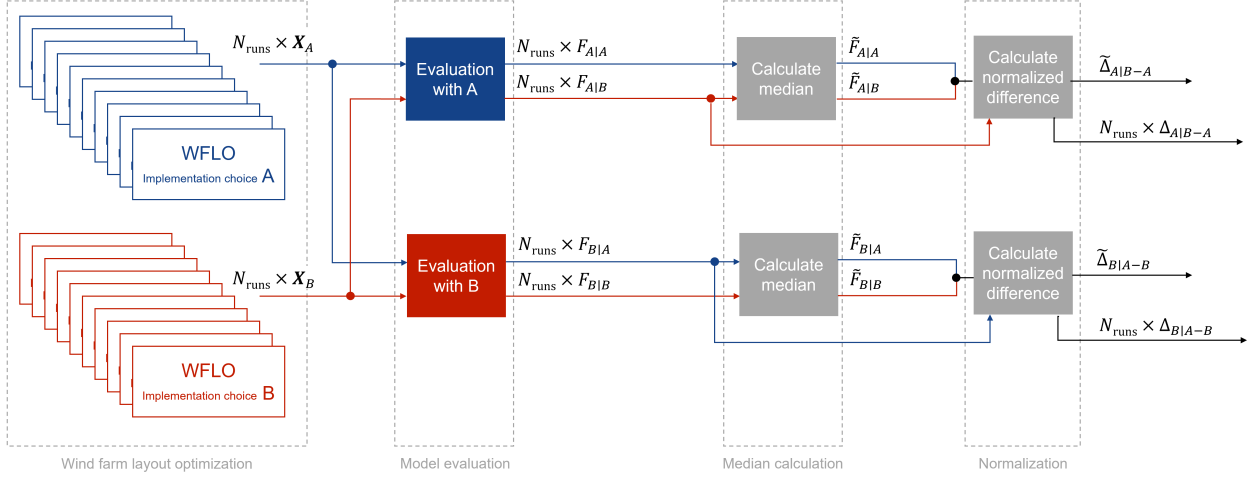


Figure 1. Calculation steps for the layout comparison based on performance metrics, for two implementation choices A and B .

As a result of this process, the metrics $\tilde{\Delta}_{A|B-A}$ and $\tilde{\Delta}_{B|A-B}$ are obtained, each providing a single measure of the impact of the implementation choices based on performance metrics. They quantify the discrepancy resulting from using an alternative evaluation choice compared with the condition in which the design and evaluation choices are identical. Therefore, negative values indicate that the resulting layouts are suboptimal when the two choices do not match, which is the most common outcome. The larger their absolute values, the greater the impact that the two implementation choices have on each other.

2.3 Layout comparison based on turbine positions

To compare layouts based on turbine positions and quantify the impact of different implementation choices, we introduce a novel metric, the *effective layout distance* d_{eff} , derived from the layout field concept developed in our previous work (Baricchio et al., 2026b). Here, we briefly summarize the approach to afterwards introduce the novel metric.

The layout field of a wind farm layout $\mathbf{X}_A$ is defined as the superposition of two-dimensional Gaussian distributions centered at the turbine positions:

$$\text{LF}_A(x, y) = \sum_{i=1}^{N_{\text{wt}}} \frac{1}{2\pi\sigma^2} \cdot \exp\left(-\frac{(x - x_{A,i})^2 + (y - y_{A,i})^2}{2\sigma^2}\right), \quad (4)$$

where (x, y) denotes a point in space, $(x_{A,i}, y_{A,i})$ represents the position of the i -th turbine in layout $\mathbf{X}_A$, and σ is the smoothness parameter of the layout field, corresponding to the standard deviation of the Gaussian distributions. Layout fields obtained from multiple layouts within a set $\mathcal{L}_A$ can be superimposed to obtain a representative layout field LF_A . Moreover, the difference between two layout fields can be computed as

$$\text{LFE}_{A-B}(x, y) = \text{LF}_A(x, y) - \text{LF}_B(x, y), \quad (5)$$



resulting in the layout field error (LFE), which provides a spatial representation of the differences between two layout sets.
 170 These differences can be quantified through the layout field integrated error (LFIE), defined as a scalar metric measuring the overall difference between $\mathcal{L}_A$ and $\mathcal{L}_B$:

$$\text{LFIE}_{A-B} = \iint_{\mathbb{R}^2} |\text{LFE}_{A-B}(x, y)| \, dx \, dy \quad . \quad (6)$$

The effective layout distance is defined as

$$d_{\text{eff}, A-B} := 2\sqrt{2}\sigma \cdot \text{erf}^{-1} \left(\frac{\text{LFIE}_{A-B}}{2 \cdot N_{\text{wt}}} \right) \quad , \quad (7)$$

175 where σ is the smoothness parameter of the layout field, and erf^{-1} is the inverse error function. The effective layout distance $d_{\text{eff}, A-B}$ expresses the difference between $\mathcal{L}_A$ and $\mathcal{L}_B$ as a scalar distance (e.g. in meters). Similarly to the LFIE, d_{eff} depends on a single hyperparameter, σ , which controls the spatial scale of the comparison, with smaller values emphasizing local differences and larger values capturing broader trends.

To motivate its definition, consider the simplified case in which both layout sets $\mathcal{L}_A$ and $\mathcal{L}_B$ contain a single layout with one
 180 turbine, with coordinates (x_A, y_A) and (x_B, y_B) , respectively. In this case, the Euclidean distance between the two turbines,

$$d^{(2T)} = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2} \quad , \quad (8)$$

can be directly related to the corresponding layout field integrated error $\text{LFIE}_{A-B}^{(2T)}$ as

$$d^{(2T)} = 2\sqrt{2}\sigma \cdot \text{erf}^{-1} \left(\frac{\text{LFIE}_{A-B}^{(2T)}}{2 \cdot N_{\text{wt}}} \right) \quad , \quad (9)$$

with $N_{\text{wt}} = 1$ in this case. This relation holds for any $\sigma > 0$. Therefore, to define the effective layout distance, the relation in
 185 Eq. (9) is extended to the general case of multiple layouts with multiple turbines.

The effective layout distance enables a direct comparison of layout sets based on turbine positions. Analogous to Sect. 2.2, we describe the procedure used to compare two sets of layouts $\mathcal{L}_A$ and $\mathcal{L}_B$, obtained with implementation choices A and B . The method consists of three steps, also illustrated in Fig. 2:

1. *Wind farm layout optimization*: This step is identical to that described for the comparison based on performance metrics
 190 and results in two sets of layouts, $\mathcal{L}_A$ and $\mathcal{L}_B$.
2. *Layout field error calculation*: The layout field error $\text{LFE}_{A-B}(x, y)$ is computed between $\mathcal{L}_A$ and $\mathcal{L}_B$. First, the layout fields of all individual layouts in $\mathcal{L}_A$ and $\mathcal{L}_B$ are computed. These are then superimposed within each set to obtain two aggregate layout fields, $\text{LF}_A(x, y)$ and $\text{LF}_B(x, y)$, which are subsequently subtracted. The resulting LFE map provides a qualitative assessment of the differences induced by the implementation choices.
- 195 3. *Integration*: The layout field error is integrated to obtain a scalar value, LFIE_{A-B} , which is then converted into the effective layout distance using Eq. (7), yielding $d_{\text{eff}, A-B}$.

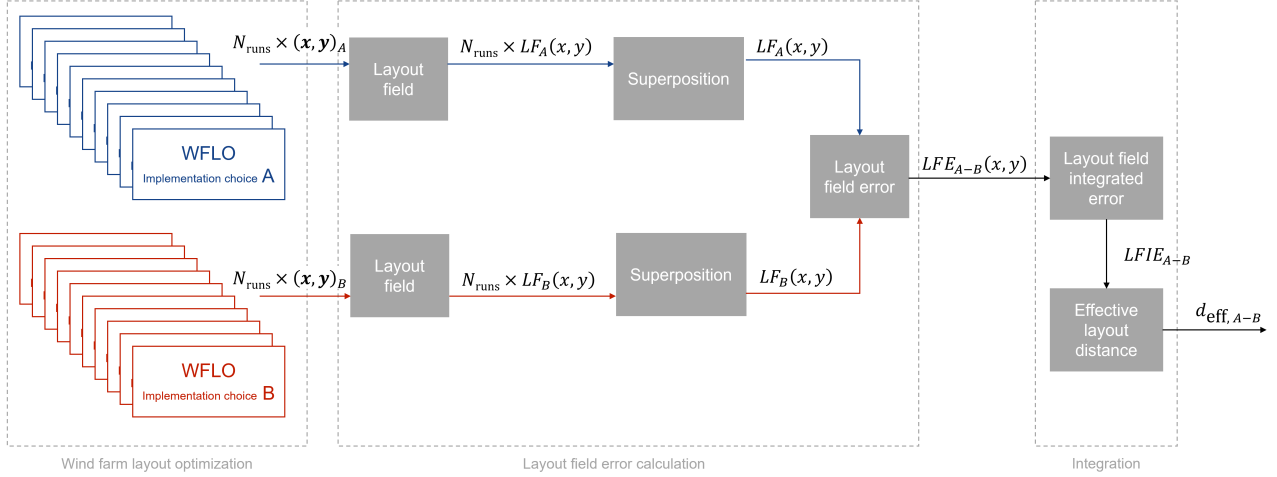


Figure 2. Calculation steps for the layout comparison based on turbine positions, for two implementation choices A and B .

Similarly to the process described in Sect. 2.2, this process yields a scalar metric $d_{\text{eff}, A-B}$ quantifying the difference between the two implementation choices. In contrast to the performance-metric-based method, this difference is assessed directly from the turbine positions, without requiring an evaluation phase. Only one value is therefore obtained, since the effective layout distance is symmetric by definition, i.e. $d_{\text{eff}, A-B} = d_{\text{eff}, B-A}$.

To provide an interpretation of comparisons between layouts obtained with different implementation choices, we introduce a classification framework to assess the impact of a given choice. This classification is based on a relative comparison with the referent implementation choice. The objective is to determine whether alternative implementation choices lead to meaningful deviations with respect to this referent.

To define the classification, two reference values REFvsREF and selfREF are defined to quantify the differences between layouts obtained from independent optimization runs using the same referent implementation choice. As discussed in Sect. 1, these differences arise from the multimodality of the WFLO problem and the stochasticity of the optimization algorithms.

The reference value REFvsREF measures the variability between two sets of layouts and is defined as

$$\text{REFvsREF} := d_{\text{eff}, \text{REF}-\text{REF}'} \quad , \quad (10)$$

namely as the effective layout distance between two independent sets of N_{runs} layouts, $\mathcal{L}_{\text{REF}}$ and $\mathcal{L}_{\text{REF}'}$, both generated using the referent implementation choice.

The reference value selfREF captures the variability between individual optimization runs and is defined as

$$\text{selfREF} := \text{median of } \{d_{\text{eff}, \text{REF}_i-\text{REF}} : i \in [1, N_{\text{runs}}]\} \quad , \quad (11)$$

with $d_{\text{eff}, \text{REF}_i-\text{REF}}$ indicating the effective layout distance between a single layout i from $\mathcal{L}_{\text{REF}}$ and all N_{runs} layouts from $\mathcal{L}_{\text{REF}}$.



Since the solutions of the WFLO problem are known to be sensitive to each optimization run, $\text{selfREF} > \text{REFvsREF}$ is expected, namely indicating that such sensitivity smoothens out when entire sets are compared instead of individual layouts. High values of both metrics indicate substantial variability among layouts obtained from independent runs, whereas values of zero indicate that the same layout is obtained in each run.

220 Based on these reference values, the impact of a given implementation choice A can be classified as follows:

- *Low impact*: $d_{\text{eff}, A-\text{REF}} < \text{REFvsREF}$. This condition indicates that the differences between layouts are smaller than the variability inherent to repeated runs of the referent case, and are therefore negligible.
- *Moderate impact*: $\text{REFvsREF} < d_{\text{eff}, A-\text{REF}} < \text{selfREF}$. This condition indicates that differences become apparent only when comparing multiple layouts, while remaining within the broader variability of the referent set.
- 225 – *High impact*: $d_{\text{eff}, A-\text{REF}} > \text{selfREF}$. This condition indicates that differences are sufficiently pronounced to be detected even when comparing individual layouts.

It can be observed that such a classification cannot be defined using the performance-metric-based comparison method described in Sect. 2.2. First, ambiguity would arise in the evaluation step. The comparison between A and REF would yield two different values rather than a single value because $\tilde{\Delta}_{\text{REF}|A-\text{REF}}$ is not symmetric with respect to A and REF. Therefore, A 230 could not be classified into a single category. Second, ambiguity would arise when defining selfREF and REFvsREF using the metrics defined in Eq. (2) and (3). Specifically, since REFvsREF requires two independent sets obtained using the same REF implementation choice, it would be unclear which set should be used during the normalization step. Moreover, constructing selfREF from the N_{runs} values of $\Delta_{\text{REF}|\text{REF}-\text{REF}}$ would, by definition, always yield a value of zero.

3 Case studies

235 This section describes the implementation choices compared using our proposed methods and the wind farm characteristics used to define selected scenarios.

3.1 Selected implementation choices

To define the implementation choices considered in this study, nine categories are introduced, each representing a different aspect of the WFLO problem. For each category, a set of alternative options is selected and compared. In the remainder of 240 the paper, we refer to each implementation choice using an abbreviation, e.g. "wdm-JENS" for the Jensen wake deficit model. All implementation choices are described in Table 1, alongside their abbreviations. The selection criteria are based on both commonly adopted approaches in WFLO studies and recent developments in the field. When not explicitly specified, the default parameters of the corresponding software tool or model are adopted for each implementation choice. All models are adopted within the software package PyWake (version 2.6.11) (Pedersen et al., 2023).



Table 1: Implementation choices considered in this study.

Category	Implementation choice	Abbreviation	Description
Wake deficit model	TurbOPark	REF	Model developed by Nygaard (2022).
	Jensen	wdm-JENS	Model developed by Jensen (1983).
	Bastankhah	wdm-BAST	Model developed by Bastankhah and Porté-Agel (2014).
	Zong	wdm-ZONG	Model developed by Zong and Porté-Agel (2020) and coupled with the Frandsen turbulence model (Frandsen, 2007).
Wake superposition model	Squared sum	REF	-
	Linear sum	wsm-LSUM	-
Rotor average model	Gaussian overlap	REF	Model described by Nygaard (2022).
	Rotor center	ram-RCEN	Evaluation of the wind speed at the rotor center.
	Circular Gauss integr.	ram-CGI	Integration using 7 points (Abramowitz and Stegun, 1970).
Wind direction discretization	Bin size: 2°	REF	-
	Bin size: 5°	wdd-5DEG	-
	Bin size: 10°	wdd-10DEG	-
Wind speed discretization	Bin size: 1 ms ⁻¹	REF	-
	Only one value	wsd-1VAL	Only a wind speed of 8 ms ⁻¹ is simulated.
Wind resource data	Heterogeneous	REF	Spatially varying Weibull scale and shape parameters (only wind speed is affected, not wind direction).
	Homogeneous	wrd-HOM	Constant Weibull scale and shape parameters.
Objective function choice	AEP	REF	Calculated using PyWake (Pedersen et al., 2023).
	LCOE	obj-LCOE	Calculated using DETECT (Kainz et al., 2024). Considered negative in sign for consistency with the WFLO formulation based on maximization.

Continued on next page



Table 1 – continued from previous page

Category	Implementation choice	Abbreviation	Description
	Revenues	obj-REV	The power production of each flow condition is weighted based on the corresponding electricity price.
	Cost of valued energy (COVE)	obj-COVE	Calculated as defined by Simpson et al. (2020) using DETECT (Kainz et al., 2024). Considered negative in sign for consistency with the WFLO formulation based on maximization.
Wind farm co-design	Only layout	REF	-
	Layout and yaw angles	cod-YAW	The yaw angles for wake steering are approximated during the AEP calculation using the “Exponential corrected” relation from (Baricchio et al., 2024).
	Layout and cables	cod-CAB	Nested optimization of the cable connections using OptiWindNet (Souza de Alencar et al., 2025), considering only the impact on the AEP in terms of electrical and availability losses through DETECT (Kainz et al., 2024).
Optimization algorithm	LO-GA	REF	Layout optimization genetic algorithm (LO-GA) developed by Baricchio et al. (2024), with a population size of 350 and 800 generations.
	SGD	opt-SGD	Stochastic gradient descend (SGD) algorithm developed by Quick et al. (2023), with a learning rate of half turbine diameter and a maximum number of iteration of 3000.
	LO-GA fast	opt-SGD	LO-GA with reduced values for the population size (48) and generations (50).

245 The referent WFLO problem is defined in this study by selecting a specific implementation choice for each category, labeled as REF in Table 1. An alternative WFLO implementation is then generated by modifying only one category at a time with respect to the referent. A full combinatorial exploration of all possible options across categories would result in a prohibitive number of WFLO cases, making the study intractable. Consequently, direct comparisons can only be made between options

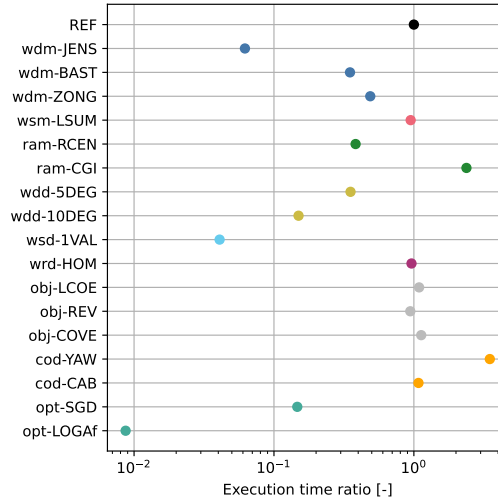


Figure 3. Computational time of each implementation choice relative to the referent. All data are based on the execution using parallel processing across 48 CPUs, except opt-SGD, which uses sequential processing on a single CPU.

within the same category and with respect to the referent. In addition, this study reports results from indirect comparisons, in which two WFLO implementations differ in their implementation choices across more than one category.

The definition of the referent follows the rationale of Thomson and Zaaier (2022), combining widely adopted implementation choices with the additional requirement that the referent configuration must support the evaluation of all alternative options across categories. For example, this requirement necessitates the use of a gradient-free optimization algorithm, allowing the inclusion of objective functions such as LCOE, for which gradient information is not readily available.

To motivate the consideration of alternative implementation choices, Fig. 3 reports the computational time associated with each option, highlighting the potentially large impact of these choices on the computational cost of WFLO. For instance, coarser discretization of wind speed or wind direction can significantly reduce the required computational time, raising the question of whether this reduction comes at the cost of meaningful differences in the resulting layout compared to the referent.

3.2 Wind farm characteristics

The impact of the different implementation choices is assessed using an offshore wind farm in the Netherlands comprising 69 IEA 22 MW reference turbines (Zahle et al., 2024), with a rotor diameter (D) of 283.2m and a hub height of 170m. This farm is representative of modern offshore wind farms and is based on a scaled version of the Hollandse Kust Noord (HKN) wind farm. The designated farm area, denoted as Ω in Eq. (1) and illustrated in Fig. 4, is obtained by scaling the original HKN boundaries while preserving its original power density of 6 W m^{-2} . The minimum distance constraint between turbines, denoted as $d_{\min}$ in Eq. (1), is set to $3.5 D$.

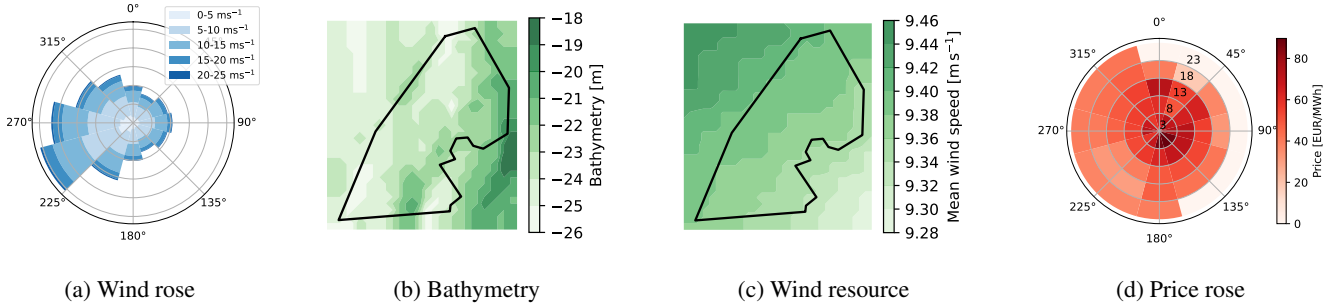


Figure 4. Baseline case study characteristics.

The wind resource for the site is obtained from modeled wind data at a 1 km spatial resolution and 16 wind direction sectors corresponding to the HKN site (Vortex FDC, 2024; Gebraad et al., 2024). The data are scaled to the turbine hub height using a power-law exponent of 0.1, which is representative of offshore environments. The resulting wind resource is visualized in Fig. 4 through the wind rose and the heterogeneous map showing the average wind speed across the domain. The latter highlights slightly higher wind speeds in the top-left region of the domain, corresponding to areas farther from the coast.

Figure 4 also shows the bathymetry profile of the considered site (Gebraad et al., 2024), which is scaled to preserve the relative spatial characteristics of the original HKN site. The water depth varies by less than 8 m across the domain, with a relatively shallow region in the central-bottom area.

The electricity prices used to compute metrics such as revenues and cost of valued energy are also shown in Fig. 4. These data are obtained from the energy system model Balmorel (Wiese et al., 2018) for the 2030 price forecast and capture the correlation between electricity prices and the local wind conditions at the site, as described by Gebraad et al. (2024). For this study, the time-series data are post-processed into statistical representations through a binning and interpolation procedure. The resulting price rose, shown in Fig. 4, specifies the electricity price for each combination of wind speed and wind direction.

The wind farm characteristics presented above define the baseline case study, which is used for most of the analyses in this study. In addition, three variants are considered to assess the sensitivity of the results to the site characteristics: low power density, high power density, and high bathymetry difference. These variants differ from the baseline case either in the farm power density or in the bathymetry profile.

The low and high power density cases are obtained by modifying the site power density to 3 W m^{-2} and 18 W m^{-2} , respectively. This is achieved by changing the scaling factor used to define the perimeter of the designated area Ω . The heterogeneous bathymetry and wind resource distributions are scaled accordingly to preserve the relative spatial characteristics of the original site. In addition, for the high power density case, the minimum turbine spacing constraint $d_{\min}$ is reduced to $2D$ to investigate the effect of denser turbine clusters under a less restrictive spacing constraint, as considered, for example, by Stanley et al. (2023).

Conversely, the high bathymetry difference case is obtained by scaling the bathymetry values to increase the variation in water depth across the domain. Specifically, the bathymetry values are scaled to cover the range $[10 \text{ m}, 60 \text{ m}]$ across all points

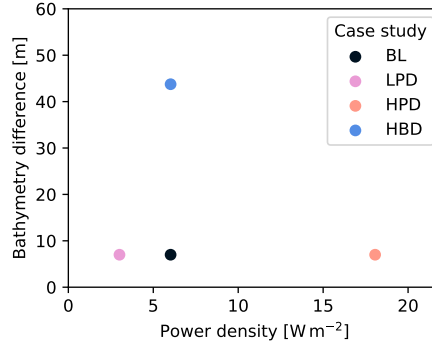


Figure 5. Characteristics of the case studies in terms of bathymetry difference and power density.

in the map shown in Fig. 4, which represents a reasonable range for bottom-fixed turbines (Kainz et al., 2024). This results in a maximum water depth difference of approximately 44m within Ω . The scaling procedure preserves the overall bathymetry profile shown in Fig. 4, while modifying only the magnitude of its spatial variations.

The characteristics of all case studies introduced in this section are summarized in Fig. 5.

295 4 Results

This section presents the results of the study, identifying which implementation choices have a significant impact on the WFLO. First, the analysis is performed using the state-of-the-art comparison approach based on performance metrics. Then, the analysis is repeated by comparing layouts based on turbine positions using the layout field error and effective layout distance concepts.

These analyses are based on 10 optimized layouts for each implementation choice, obtained from $N_{\text{runs}} = 10$ independent
 300 optimizations for each case. The sensitivity of the results to the choice of N_{runs} is discussed in Appendix B.

4.1 Layout comparison based on performance metrics

We first examine the impact of the implementation choices based on performance metrics, following the method described in Sect. 2.2.

The results are first presented using the metric $\Delta_{A|B-A}$ defined in Eq. (2). This metric is used to compare a layout optimized
 305 using implementation choice j as the design choice with the referent, used as both the evaluation choice and the basis for normalization. The corresponding metric is therefore $\Delta_{\text{REF}|j-\text{REF}}$. The results are shown in Fig. 6, where boxplots represent the distribution of $\Delta_{\text{REF}|j-\text{REF}}$ across the N_{runs} realizations for each implementation choice j , indicated on the x-axis.

First, it can be observed that most values are negative. This behavior results from the selected normalization approach, in which the normalization term corresponds to the case where the design and evaluation choices are identical. Consequently, the
 310 impact of an alternative implementation choice j relative to the referent is reflected by a reduction in performance with respect

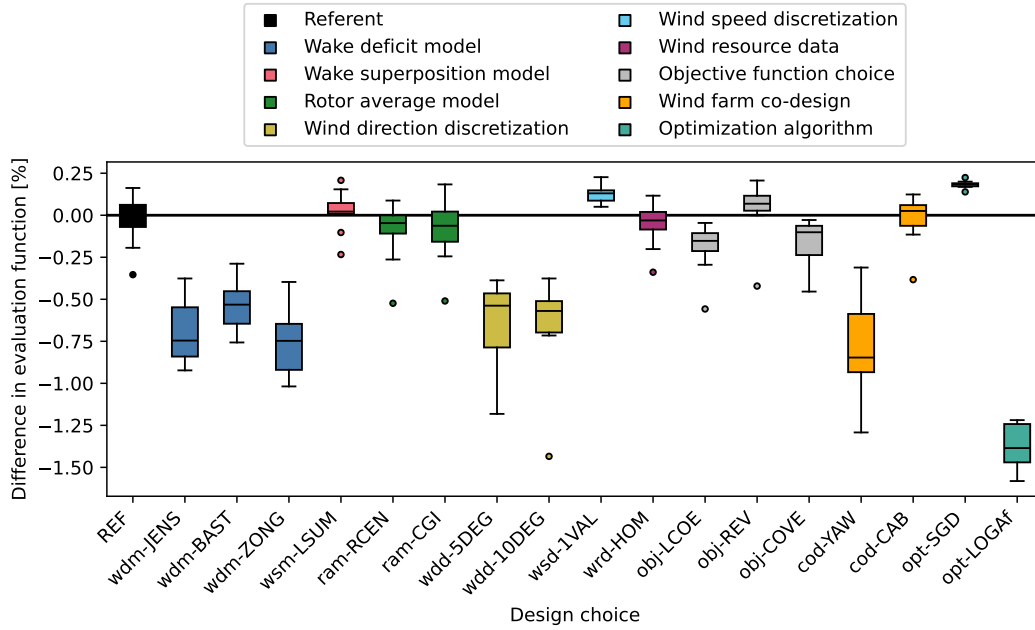


Figure 6. Comparison between the different implementation choices, expressed in terms of $\Delta_{\text{REF}|j-\text{REF}}$, with j indicated on the x-axis. Each boxplot results from 10 layouts obtained from independent optimization runs of the same implementation choice. Performance metrics that are minimized, such as LCOE, are expressed with a negative sign (e.g., $-\text{LCOE}$); i.e. an increase in the corresponding metric value represents an improvement.

to the referent. More negative values, therefore, indicate a larger deviation from the referent and, consequently, a greater impact of the implementation choice.

The results in Fig. 6 show that the largest deviation from the referent is obtained when reduced hyperparameter values are used for the LO-GA optimization algorithm (opt-LOGAf). Large deviations are also observed when an alternative wake deficit model is used instead of TurbOPark (wdm-JENS, wdm-BAST, wdm-ZONG), when a coarser wind direction resolution is used (wdd-5DEG, wdd-10DEG), and when the control co-design approach is adopted (cod-YAW). According to this comparison method, these implementation choices therefore have the largest impact on the WFLO results. In contrast, most of the other implementation choices produce values that largely fall within the range of the referent, indicating a limited impact based on performance metrics.

Another relevant observation is the presence of positive values for some implementation choices, which may appear counterintuitive given the normalization approach. This behavior is related to the performance of the optimization algorithm and its interaction with the selected design choices. This effect is most straightforwardly observed when the SGD algorithm is used (opt-SGD). In this case, the positive values indicate that SGD achieves slightly better performance than the reference LO-GA under these conditions. However, the difference is negligible, and the values associated with the best-performing solutions almost coincide. A less intuitive behavior is observed when only one wind speed value is considered (wsd-1VAL). This condition



avoids the simulation of edge cases associated with the transition between operating regimes for clusters of turbines around the cut-in, rated, and cut-out wind speeds. This simplifies the design space and, as a result, facilitates convergence towards improved solutions, likely by reducing the multimodality of the problem.

In Fig. 6, the co-design category shows negative deviations. This results from evaluating the resulting layouts using the evaluation function associated with the referent implementation choice rather than with the co-design choice. For example, the values for control co-design (cod-YAW) correspond to AEP evaluated without wake steering, even though wake steering is included during the design phase. Therefore, these results do not quantify the benefits of co-design. Instead, they quantify the impact of introducing co-design on the resulting WFLO solutions relative to the referent.

Thus far, each alternative implementation choice has been compared only with respect to the referent. We now extend this analysis by considering cross-comparisons between implementation choices within the same category (direct comparison) and, for completeness, across different categories (indirect comparison).

The results are presented as a matrix plot in Fig. 7, using the metric $\tilde{\Delta}_{A|B-A}$ defined in Eq. (3). Specifically, each cell reports the value of $\tilde{\Delta}_{i|j-i}$ for evaluation choice i on the y-axis and design choice j on the x-axis. The diagonal entries, therefore, correspond to cases in which the same implementation choice is used for both design and evaluation. These entries are equal to zero because they serve as the normalization terms for the corresponding rows, following the definitions in Eq. (2)-(3). Consequently, the first row of the matrix contains the median values shown in the boxplots of Fig. 6.

A few rows, corresponding to evaluation choices, are left empty because the resulting comparisons would not provide meaningful information for either of two reasons. First, the evaluation function may rely on an approximation of the referent implementation choice and is therefore inherently less accurate than the referent. This applies to the wind direction discretization, wind speed discretization, and wind resource data categories. Second, the evaluation function may be identical to that of the referent, as is the case when comparing different optimization algorithms. Including these comparisons would be redundant with the first row and would distort the normalized values because of the larger deviations observed for the implementation choice LO-GA with reduced hyperparameter values (opt-LOGAf).

The results in Fig. 7 confirm the trends observed in the comparison with the referent, with larger deviations concentrated in the columns associated with the implementation choices previously identified as having a larger impact. However, the direct comparison between the different wake deficit models reveals smaller deviations when the alternative models, Jensen, Bastankhah, and Zong, are compared with each other. This suggests that the referent choice for this category, TurbOPark, differs more substantially from the other models, which exhibit greater consistency among themselves.

Another relevant observation concerns the row associated with control co-design (cod-YAW). Several positive values are obtained in this row, which may appear counterintuitive given the adopted normalization. This indicates that layouts obtained using this implementation choice can have lower performance, even when evaluated with the same implementation choice. This behavior is related to the effect of this implementation choice on the optimization process. Specifically, including wake steering in the objective function increases the complexity of the design space, potentially introducing additional local optima that hinder convergence.

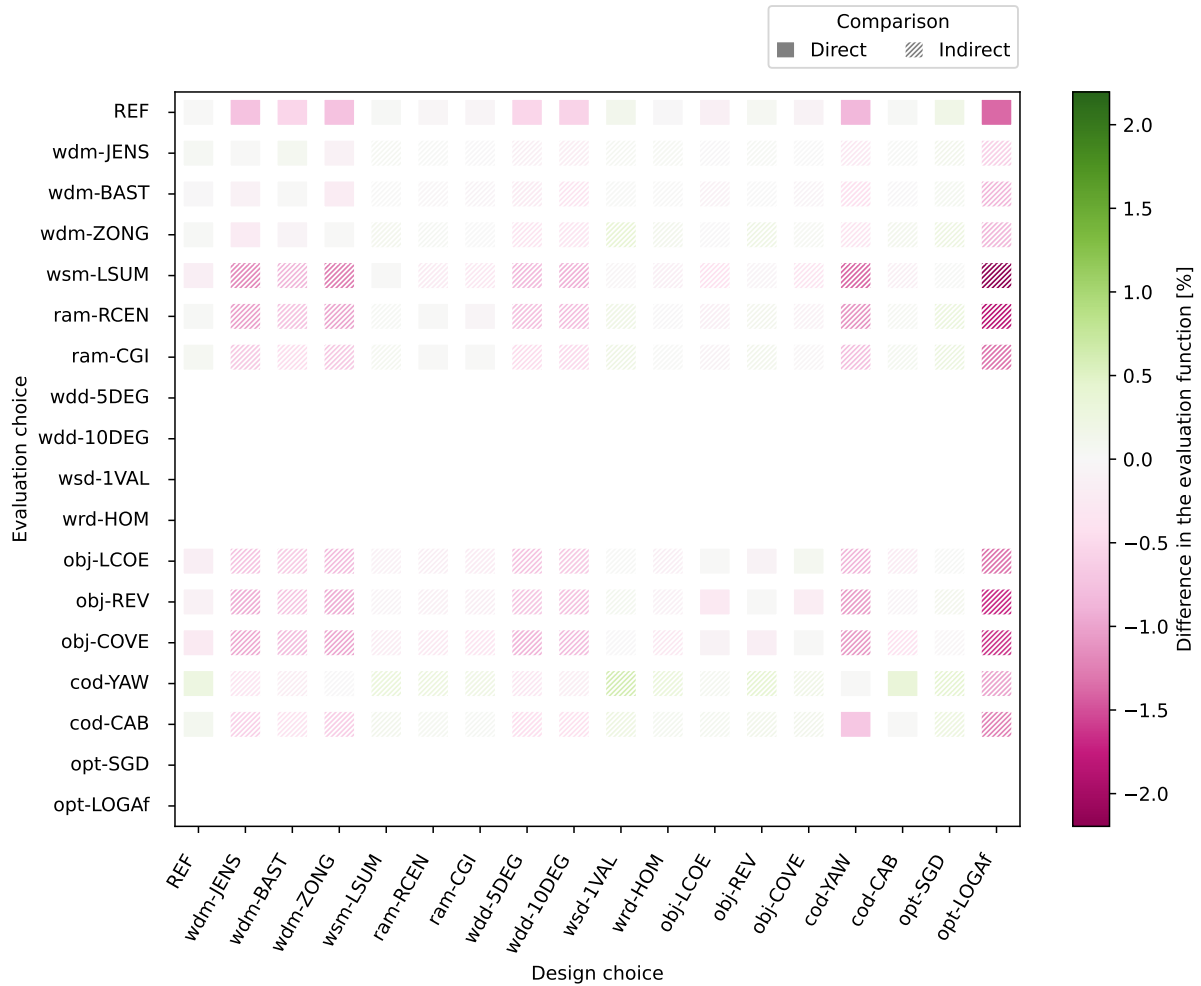


Figure 7. Comparison of implementation choices based on the resulting performance metrics. Each cell reports the value $\tilde{\Delta}_{i|j-i}$, with j indicating the design choice shown on the x-axis and i the evaluation choice shown on the y-axis. Values corresponding to irrelevant evaluation choices are omitted, and it is indicated whether each comparison between choices is direct or indirect. Performance metrics that are minimized, such as LCOE, are expressed with a negative sign (e.g., $-LCOE$); i.e. an increase in the corresponding metric value represents an improvement.

360 To assess whether the observations derived from the baseline scenario can be extended to a broader range of wind farm characteristics, the analysis is repeated for the three additional cases introduced in Sect. 3.2. The results focus on the impact of the different implementation choices relative to the referent, following the same procedure used to construct the first row of the matrix in Fig. 7, which is here replicated for each scenario. The outcomes are presented in Fig. 8 using a similar visualization, where each column corresponds to a specific implementation choice adopted during the design phase, and each row represents

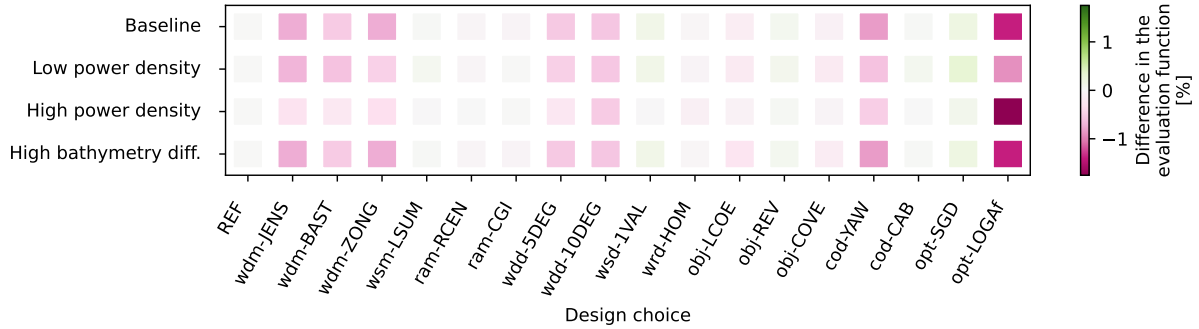


Figure 8. Comparison of the implementation choices evaluated with the referent performance metric for different wind farm characteristics (shown on the y-axis). Each cell reports the value $\tilde{\Delta}_{\text{REF}|j-\text{REF}}$, with j indicating the design choice shown on the x-axis.

a different scenario. For each scenario, the referent implementation choice is consistently applied during the evaluation across all four cases.

The observations from the baseline scenario remain largely unchanged across all scenarios. Larger deviations from the referent are consistently observed for reduced LO-GA hyperparameter values, followed by alternative wake models, coarser wind direction discretizations, and the control co-design approach. The impact of the different wake models is slightly lower in the high power density scenario, whereas reduced LO-GA hyperparameter values have an even larger impact in this case.

4.2 Layout comparison based on turbine positions

In this section, we compare the impact of different implementation choices based on turbine positions using two metrics: the layout field error and the effective layout distance, which provide qualitative and quantitative comparisons, respectively. The analyses are performed using a smoothness parameter of $\sigma = 5, D$, consistent with the criterion based on the average minimum turbine spacing reported in our previous work (Baricchio et al., 2026b). The sensitivity to this parameter is further analyzed in Appendix A.

Figure 9 shows the LFE obtained for different wake deficit models as an example. The LFE is computed for pairwise comparisons among the four implementation choices: TurbOPark (REF), Jensen (wdm-JENS), Bastankhah (wdm-BAST), and Zong (wdm-ZONG). The four alternatives result in six pairwise comparisons, each represented by a separate map. The LFE maps indicate that, when TurbOPark (REF) is used, turbines tend to be positioned closer to the domain boundaries than when the other models are used. In contrast, the differences among the Jensen, Bastankhah, and Zong models are less pronounced and are not readily distinguishable in the LFE maps.

The effective layout distance, d_{eff} , is computed for all implementation choices described in Sect. 3.1 with respect to the referent, following a procedure analogous to that used for the performance-metric-based comparison. The results are presented as a histogram in Fig. 10, with the selfREF and REFvsREF values highlighted as thresholds for the impact classification. These thresholds enable the rigorous classification of each implementation choice as having low, moderate, or high impact, as defined

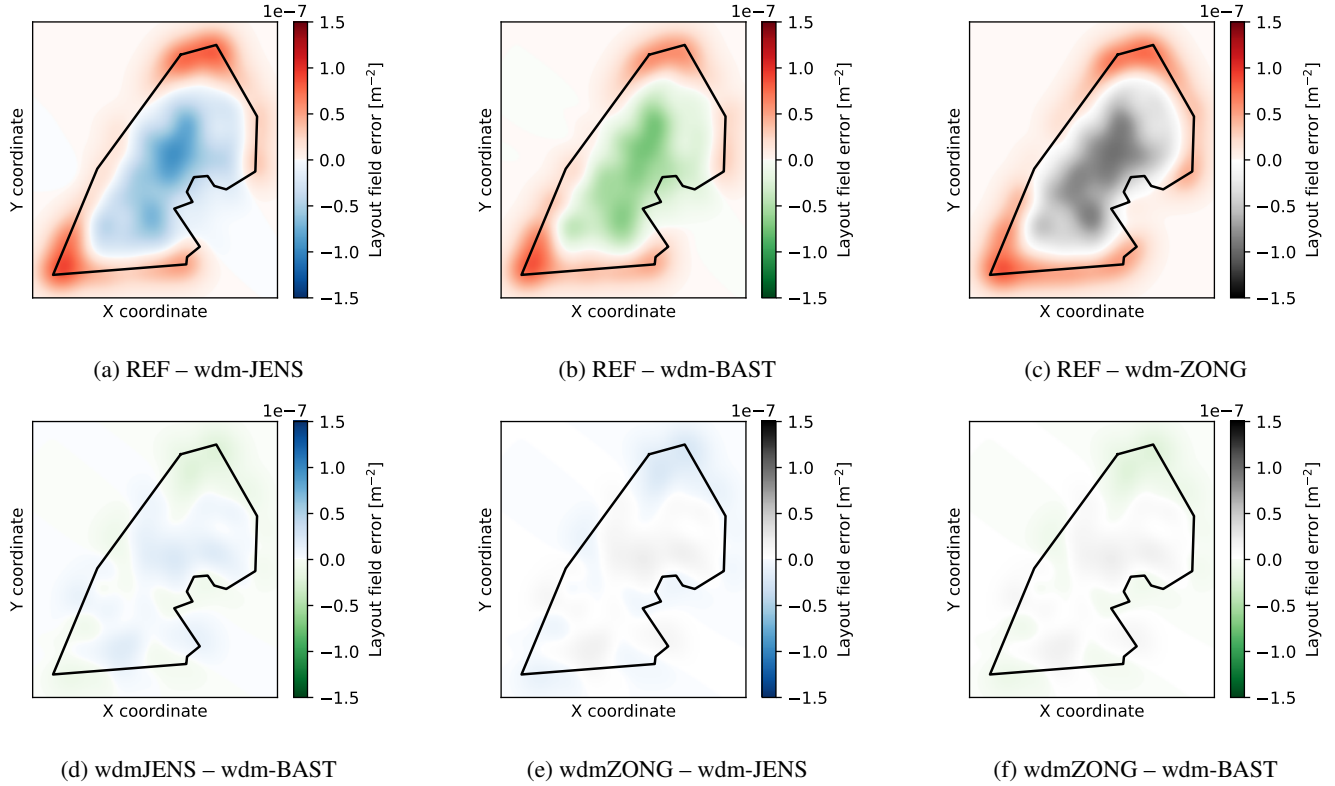


Figure 9. Layout field error between the sets of layouts obtained with different wake deficit models. A different color is used to indicate each model: TurbOPark (red), Jensen (blue), Bastankhah (green), and Zong (grey). The compared models are specified in each subplot, with the first term indicating the positive values in the LFE and the second indicating the negative ones.

in Sect. 2.3. In contrast to Fig. 6, no distribution is shown because each d_{eff} value already summarizes the differences between the corresponding sets of layouts.

The largest d_{eff} value is obtained when reduced hyperparameter values are adopted for the optimization algorithm (opt-LOGAf) to accelerate the optimization process, followed by the use of alternative wake deficit models (wdm-JENS, wdm-BAST, wdm-ZONG). These implementation choices, therefore, result in layouts that differ substantially from those obtained with the referent.

Although its d_{eff} value is only slightly above the threshold for moderate impact, the use of cost of valued energy as the objective (obj-COVE) is classified as a high-impact choice. This contrasts slightly with the previous analysis, where only a minor deviation in performance metrics was observed. The result suggests that using COVE as the objective drives the turbines towards specific regions of the domain, such as areas with lower bathymetry, resulting in a layout that differs from the referent while achieving similar performance. This behavior can be attributed to the multimodality of the WFLO problem, where substantially different layouts can yield similar performance metric values. Therefore, the high impact of obj-COVE in terms of turbine positions does not necessarily translate into a corresponding impact on performance. This effect is less pronounced

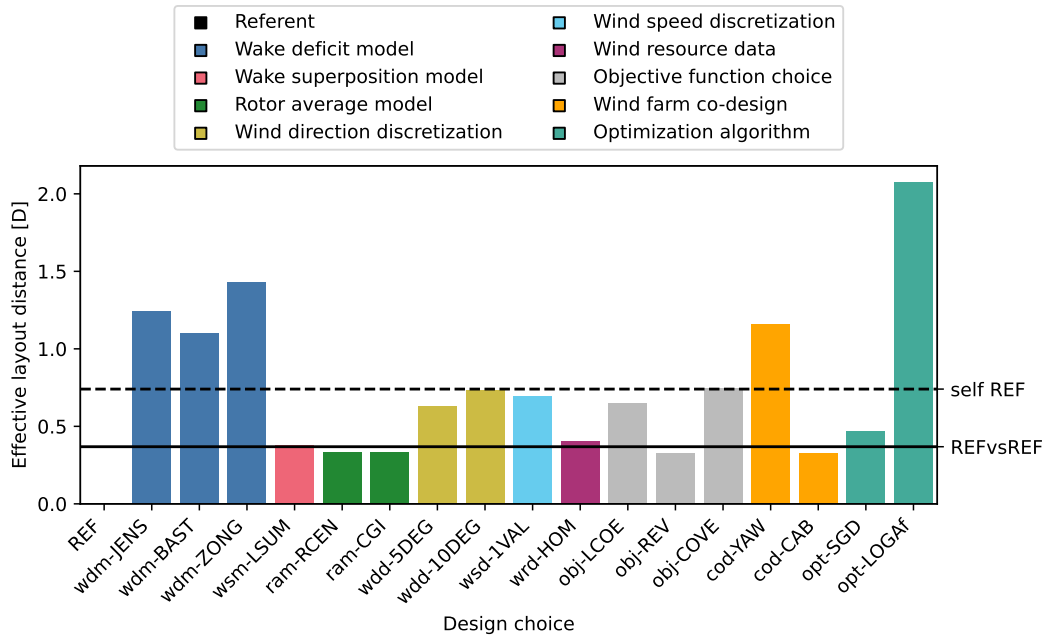


Figure 10. Effective layout distance between the sets of layouts obtained with the different implementation choices and the referent. The values selfREF and REFvsREF are reported to enable the classification into low-, moderate-, and high-impact.

or absent for the other objectives, whose impact ranges from moderate for the levelized cost of energy (obj-LCOE) to low for revenues (obj-REV).

For wind direction discretization (wdd-5DEG and wdd-10DEG), the impact on the layouts is classified as moderate, whereas a larger difference was observed on the performance metrics. This discrepancy arises because coarser wind direction discretization can introduce layout variations associated with turbine alignment along directions that are not explicitly simulated. These variations are typically local and relatively small and are therefore attenuated by the effective layout distance. Their influence, however, remains more pronounced when the performance metrics are compared. Conversely, wind speed discretization (wsd-1VAL) shows the opposite behavior, with a moderate impact according to d_{eff} , compared with the minor deviations observed in the performance-metric-based comparison.

The remaining implementation choices have d_{eff} values close to the boundary between low and moderate impact. These include the wake superposition models (wsm-LSUM), rotor averaging models (ram-RCEN, ram-CGI), wind resource data (wrd-HOM), cable co-design (cod-CAB), and an alternative optimization algorithm (opt-SGD). Therefore, these choices have a limited impact on the resulting layouts compared with the referent.

Analogous to the performance-metric-based analysis, the comparison is extended to include cross-comparisons among all implementation choices, covering both direct and indirect cases. The results are presented in the matrix plot in Fig. 11, where only the impact classification into low, moderate, and high is reported rather than the corresponding d_{eff} values. A key differ-



ence from Fig. 7 is that both axes represent design choices, as this comparison does not involve an intermediate evaluation step using a separate implementation choice. Consequently, the matrix in Fig. 11 is symmetric. The diagonal entries, corresponding to comparisons between identical sets of layouts, have $d_{\text{eff}} = 0$ by definition and are therefore omitted from the plot.

The results shown in Fig. 11 highlight the high impact of adopting alternative wake deficit models (wdm-JENS, wdm-BAST, wdm-ZONG) and using reduced hyperparameter values for the optimization algorithm (opt-LOGAf). These implementation choices have a high impact not only when compared with the referent, as shown in Fig. 10, but also when compared directly with other implementation choices through the cross-comparison.

Unlike the performance-metric-based analysis, Fig. 11 contains values for all rows, thereby also enabling cross-comparison between implementation choices that approximate the referent, such as those involving coarser wind direction discretizations. This is possible because the proposed comparison method does not require a separate evaluation step.

The cross-comparison of the wake deficit models further highlights the trend identified in the performance-metric-based analysis. Specifically, a low impact is obtained when comparing the Jensen (wdm-JENS) and Bastankhah (wdm-BAST) models, whereas a moderate impact is obtained when either of these models is compared with the Zong model (wdm-ZONG). This indicates that Jensen and Bastankhah result in similar turbine positions, while the Zong model produces larger differences. Consistent with the previous analysis, TurbOPark (REF) results in substantially different turbine positions compared with all other wake deficit models.

In contrast to the performance-metric-based analysis, the cross-comparison of the objectives considered in this study, namely levelized cost of energy (obj-LCOE), revenues (obj-REV), and cost of valued energy (obj-COVE), results in a moderate impact. This indicates that these implementation choices have a greater influence when assessed based on turbine positions than based on their resulting performance metrics.

The analysis conducted for the baseline scenario is extended to the three additional cases, namely low power density, high power density, and high bathymetry difference. An analogous methodology is adopted, focusing on the impact of the different implementation choices relative to the referent. In this context, the first row of the matrix in Fig. 11 is replicated for each of the additional cases. The corresponding results are presented in Fig. 12.

This sensitivity analysis indicates that the number of high-impact implementation choices increases for high power density sites. Compared to the baseline scenario, a high impact is also observed for coarser wind direction and speed discretization (wdd-10DEG and wsd-1VAL) and for the use of LCOE as the objective (obj-LCOE). The opposite trend is observed for low power density sites, where only the Jensen wake deficit model (wdm-JENS) and reduced LO-GA hyperparameter values have a high impact. For sites with large bathymetry variations, objectives such as LCOE, which penalize turbines located in deeper waters, also have a high impact.

These results differ visibly from the performance-metric-based analysis, where the same trends are consistently observed across all scenarios. In contrast, the effective layout distance captures how site conditions affect the impact of implementation choices. This is particularly evident for the different objective choices in the high bathymetry scenario. Comparing turbine positions can therefore reveal how the dominant drivers of the optimization shift with site conditions, an effect that may be masked by the multimodality of the WFLO problem when considering only performance metrics.

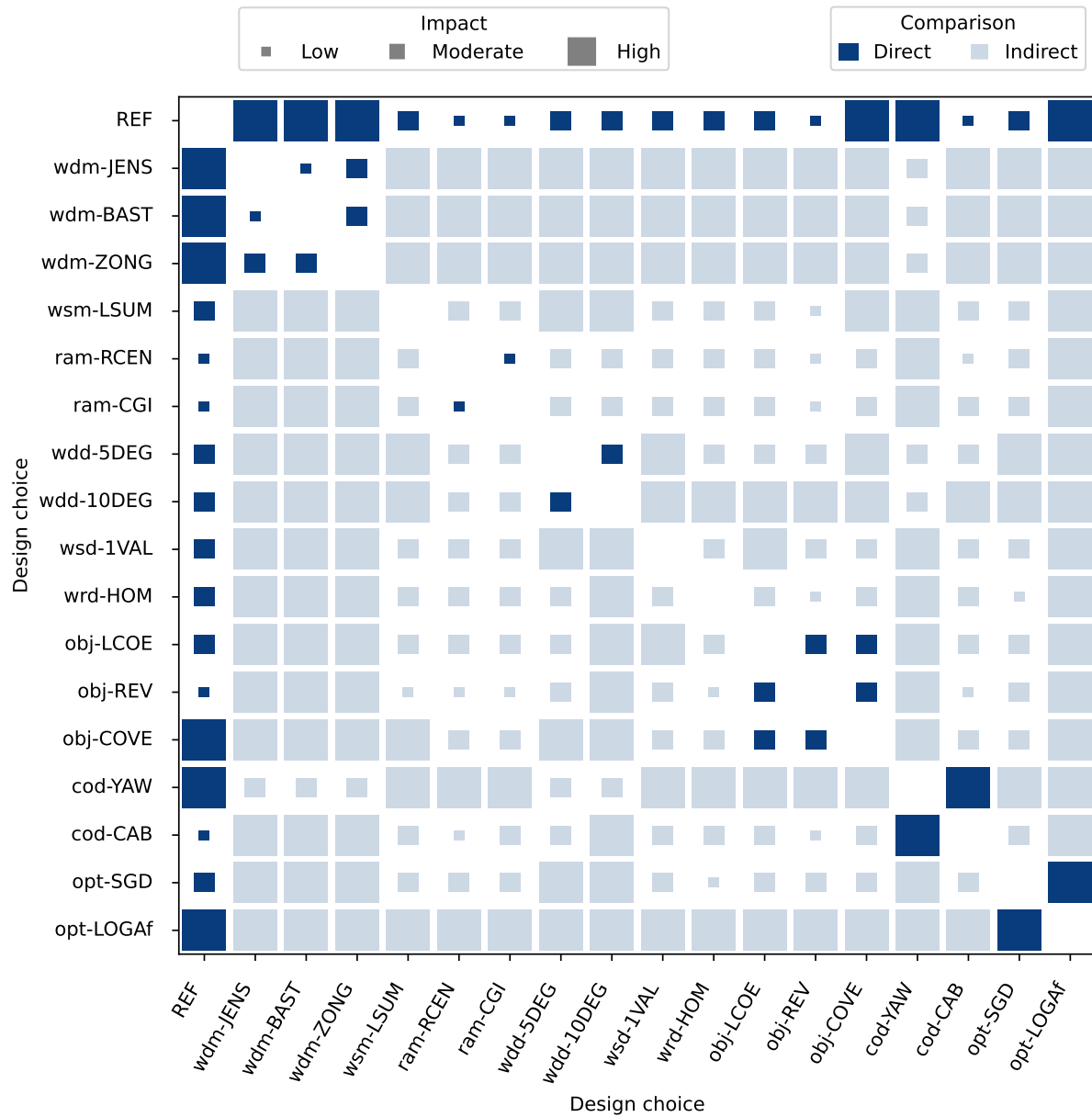


Figure 11. Comparison between the implementation choices based on the effective layout distance. Each cell corresponds to the d_{eff} between sets of layouts obtained with the implementation choices specified on the x- and y-axis. These values are classified into low-, moderate-, and high-impact. Whether the comparison between choices is direct or indirect is also specified.

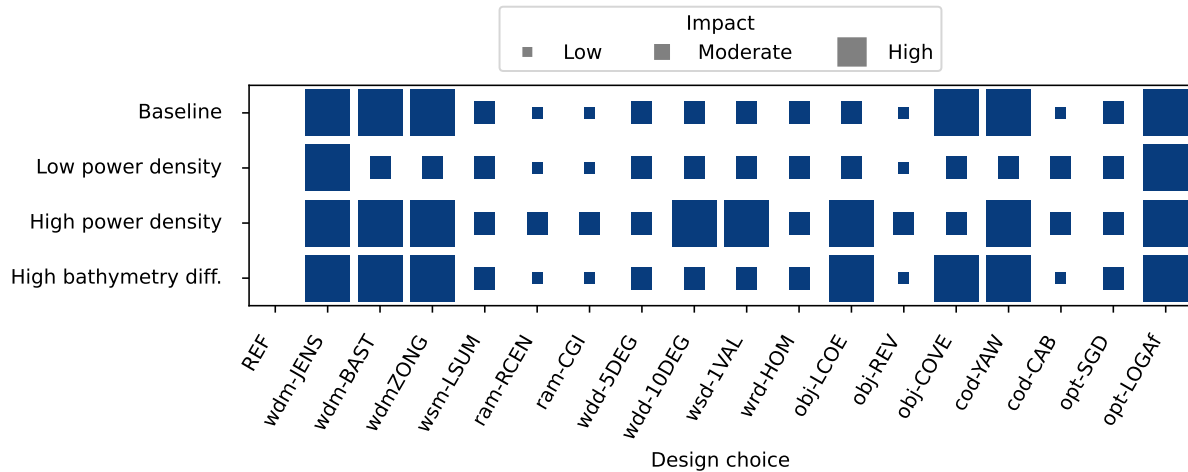


Figure 12. Comparison between the different implementation choices (shown on the x-axis) and the referent for different wind farm characteristics (shown on the y-axis) based on d_{eff} .

5 Discussion

This section discusses the results of this study, highlighting the relevance of the proposed method and the implications of the findings for wind farm layout optimization. First, the advantages of using the effective layout distance over the state-of-the-art comparison method based on performance metrics are discussed. Second, the results are analyzed to understand the practical implications of low, moderate, and high impact for the different implementation choices. Third, the limitations of the study are discussed.

5.1 Advantages of comparing layouts with the effective layout distance

The comparison of different layouts based on the effective layout distance avoids three issues that are intrinsic to the performance-metric-based method.

First, it avoids model dependence during the comparison by directly considering the primary outcome of WFLO: the turbine positions. Conversely, the performance-metric-based comparison relies on models that translate turbine positions into a representative performance value. These models are often the same as those employed during the WFLO process and are therefore typically of lower fidelity, i.e. relatively less accurate. They may also differ from the higher-fidelity models used in subsequent stages of wind farm assessment. Such discrepancies introduce additional uncertainty into the evaluation of layouts and, consequently, their comparison. An example of this issue is observed for the wind direction discretization. Although a sufficiently fine resolution is known to be required in WFLO to avoid turbine alignments along directions that are not explicitly simulated, the impact measured through d_{eff} is moderate rather than high, as might be expected from the performance-metric-based results. This difference arises because d_{eff} is less sensitive to small and localized changes in turbine positions caused by



coarse wind direction discretizations. Such local variations would likely occur anyway during the practical development of a
470 wind farm due to installation-related constraints. Conversely, d_{eff} is more sensitive to larger-scale changes, such as variations
in the regions of the domain occupied by turbines. An opposite behavior is observed for the objective choice. In some cases,
particularly for the cost of valued energy, a high impact is detected using d_{eff} despite only minor differences in performance
metrics. This can be explained by the fact that different objectives may drive the turbine positions according to distinct aspects
of the site characteristics. For example, layouts optimized to reduce costs associated with bathymetry may introduce additional
475 wake losses, leading to compensating effects that result in similar overall performance metrics. Overall, this behavior is also a
consequence of the multimodality of the WFLO problem, where significantly different layouts can achieve comparable perfor-
mance values. From a practical perspective, such differences in turbine positioning can be extremely relevant, and the proposed
 d_{eff} -based comparison enables their identification.

Second, the effective layout distance requires only one variable parameter for the comparison, namely the smoothness
480 parameter σ . Conversely, the performance-metric-based method involves several methodological choices that can affect the
comparison. One such choice is the normalization procedure. In this study, normalization is performed with respect to the case
in which the implementation choice adopted during the design phase is consistent with that used during the assessment phase.
In principle, alternative normalization strategies could be adopted, potentially leading to different conclusions.

Third, the effective layout distance intrinsically enables the comparison between two sets of layouts rather than only between
485 two individual layouts. Comparing sets of layouts allows implementation choices to be classified based on their overall impact
on turbine positions, distinguishing systematic trends from the variability between individual optimization runs arising from
the multimodality of the WFLO problem. This variability can be interpreted as a form of “noise” when identifying systematic
effects. Conversely, performance-metric-based comparisons often focus on the best-performing layout among the different
optimization runs. However, the selection of this layout depends on a specific performance model, which may itself be subject
490 to uncertainty. Moreover, when the objective is to identify systematic trends in WFLO rather than isolated best-performing
solutions, focusing on a single layout may not provide a representative assessment of the impact of an implementation choice.

5.2 What matters for layout optimization?

This study identifies the impact of different implementation choices in WFLO and classifies them as having low, moderate,
or high impact. This classification provides practical guidance on which implementation choices require careful consideration
495 when formulating a WFLO problem and where priorities should be placed when trade-offs between accuracy, complexity, and
computational cost need to be considered.

A high-impact choice indicates that adopting it instead of the referent can lead to substantial differences in the resulting
turbine positions. When such an impact arises from a simplification of the WFLO problem, it suggests that the simplification
may be too substantial and could significantly affect the optimization outcome. An example observed in this study is the use
500 of reduced hyperparameters for the optimization algorithm. Although these settings reduce the computational time required
to obtain a solution, they result in lower-quality solutions and therefore represent a simplification that should be avoided
when the quality of the optimized layout is important. Conversely, when a high-impact choice arises from different modeling



approaches, such as the wake deficit models considered in this study, there may be no established basis for identifying one approach as universally better than another. In this case, the results indicate that particular attention should be paid to the model selection. For example, additional effort could be devoted to assessing which modeling approach provides the most realistic representation of the physical processes, since the choice can have a substantial effect on the final turbine positions.

In contrast, a low-impact choice indicates that selecting it instead of the referent results in similar turbine positions. For the purpose of WFLO, the choice can therefore be guided primarily by practical considerations, such as model simplicity, availability, and computational efficiency. An example observed in this study concerns the rotor averaging model. As shown in Fig. 3, the different rotor averaging models require substantially different computational times. However, a low impact is observed for all pairwise comparisons between these choices. This indicates that selecting either the fastest model or a more accurate but computationally more expensive model results in broadly similar turbine positions. In this case, the choice can therefore be guided entirely by computational efficiency.

For implementation choices classified as having a moderate impact, the distinction between these two approaches is less clear. Such choices may warrant additional consideration, but they may also be selected based on practical considerations, depending on the specific scope of the WFLO study. The appropriate choice is therefore left to the designer.

Understanding which implementation choices can be simplified or given less attention enables wind farm designers to allocate computational resources to aspects of WFLO that are more relevant. For example, computational resources could instead be used to enable the simulation of more accurate flow models or to incorporate additional aspects of wind farm design through a co-design approach. While this study considers co-design only for two examples, the same principle could be extended to a much broader range of wind farm design aspects.

5.3 Limitations

The methodology and results presented in this study also have some limitations that should be considered when interpreting the findings.

First, the classification into low, moderate, and high impact depends on the definition of the referent. Therefore, selecting different models, datasets, or optimization algorithms as the referent could lead to different classifications. However, the referent in this study is defined based on state-of-the-art models and commonly adopted implementation choices. Consequently, we expect similar conclusions to be obtained for slight variations in its definition.

Second, the implementation choices considered in this study do not exhaust the range of possibilities available within the field of WFLO. A comprehensive assessment of all possible choices would be impractical given the large number of alternatives available in the literature. This is particularly evident for optimization algorithms, for which numerous methods specifically developed for WFLO have been proposed in recent years. Moreover, the aim of this study is not to provide an exhaustive assessment of each individual implementation choice, but rather to demonstrate the impact of choices across substantially different categories, ranging from optimization algorithms to modeling approaches. Nevertheless, the proposed methodology can be directly applied to additional implementation choices. The analysis could therefore be extended to investigate more specific aspects of WFLO, such as the sensitivity to individual parameters of the wake deficit models.



Third, the comparison based on the effective layout distance determines whether an implementation choice leads to different turbine positions, but it does not establish whether a particular approach can improve wind farm performance through WFLO. This limitation is particularly relevant in the context of co-design, where a change in the resulting layout does not in itself indicate whether the overall wind farm design is improved. In this context, the effective layout distance should therefore be regarded as a complementary assessment tool rather than as a standalone method for evaluating the benefits of a particular implementation choice.

Finally, the results presented in this study may vary to some extent depending on the selected value of σ and on the number of layouts compared for each implementation choice, i.e. N_{runs} . However, the sensitivity to these assumptions is investigated in Appendices A and B, respectively. The results of these analyses demonstrate that the main conclusions remain robust with respect to variations in these parameters.

6 Conclusions

This study quantified the impact of different implementation choices in wind farm layout optimization, identifying those with significant and negligible effects on both performance metrics and resulting turbine positions.

Two different methods were adopted to compare implementation choices. First, a conventional approach based on performance metrics was used, in line with existing literature. Second, a method based on a novel metric named effective layout distance was introduced, building on the concept of layout field error proposed in our previous work. This second approach shifts the focus from performance indicators to the primary outcome of WFLO, namely turbine positions. Furthermore, it enables a classification of implementation choices into low-, moderate-, and high-impact categories.

The results showed that, for a 69-turbine offshore wind farm, four aspects had a high impact. Most notably, the selection of optimization algorithm hyperparameters played a critical role. In addition, the choice of wake deficit model strongly affected the layouts; for example, TurbOPark clustered more turbines toward the farm perimeter compared to other models. Significant differences were also observed when the cost of valued energy is used as objective instead of more traditional metrics such as AEP. Lastly, incorporating wake steering through the geometric yaw approach in AEP calculations also influenced the resulting layouts. These findings indicate that such choices require careful consideration in WFLO, as they can lead to substantial differences in turbine positions and, consequently, have relevant practical implications.

Conversely, other implementation choices, such as wake superposition methods, rotor-averaging models, and the inclusion of electricity price information when optimizing revenue instead of AEP, led to negligible differences in the layouts. This suggests that, for these aspects, simpler and more computationally efficient models can be preferred.

These results remain largely consistent across different site conditions, with only a few discrepancies. For high power density sites, more implementation choices have a high impact, whereas the opposite is observed at lower power densities. For sites with large bathymetry variations, objectives that penalize greater water depths can also have a high impact.

Therefore, the comparison method introduced in this study can support future WFLO research by enabling new results to be assessed in a more systematic and consistent manner. This is particularly relevant when introducing new or more efficient



570 optimization algorithms or investigating the impact of specific modeling parameters. Moreover, for the implementation choices considered in this study, the results of this study provide wind farm designers with insight into which aspects of WFLO require greater attention and where computational resources can be allocated more efficiently.

Code and data availability. The software and data used in this study are available at <https://doi.org/10.5281/zenodo.23016905> under the folder examples/wflo.

575 **Appendix A: Sensitivity analysis on the smoothness parameter**

This analysis investigates the robustness of the results presented in this study with respect to the smoothness parameter σ , which represents the only variable parameter in the comparison method based on d_{eff} .

The analysis on the comparison between each implementation choice and the referent presented in Fig. 10, in Sect. 4.2, is repeated for a range of σ values. The results are shown in Fig. A1, in which each curve corresponds to a specific implementation
580 choice. Similarly to Fig. 10, also the line corresponding to selfREF and REFvsREF values are reported. To assess whether the impact of the different choices varies with σ , the relative positions of the curves should be considered, both with respect to each other and, more importantly, with respect to the selfREF and REFvsREF thresholds.

The values of selfREF and REFvsREF remain approximately constant across the investigated range of σ , with only a slight decreasing trend. By definition, larger values of d_{eff} with increasing σ indicate larger differences occurring at larger spatial
585 scales. Therefore, the decreasing trend observed for selfREF and REFvsREF indicates that the diversity within $\mathcal{L}_{\text{REF}}$ decreases when considering progressively larger-scale spatial patterns.

Some implementation choices exhibit an increasing d_{eff} with increasing σ , particularly those previously identified as having a high impact. This indicates that the differences between these layouts become more pronounced when the comparison focuses on larger-scale spatial distributions rather than local turbine-level variations. Conversely, the remaining implementation choices
590 show limited sensitivity to the value of σ .

Overall, the relative ordering of the implementation choices remains unchanged for most values of σ . The few observed crossings occur only between curves with very similar values and therefore do not affect the overall interpretation. Consequently, the conclusions drawn in this study can be considered robust with respect to the selected value of σ .

Appendix B: Sensitivity analysis on the number of layouts

595 The scope of this analysis is to assess whether ten independent optimization runs of the same case, i.e. $N_{\text{runs}} = 10$, are sufficient to draw robust conclusions on the impact of the different implementation choices using the proposed method. Therefore, this analysis focuses on the number of layouts used in each comparison.

Since increasing N_{runs} requires significant computational resources, the analysis is limited to three implementation choices: the referent, from which the selfREF and REFvsREF values are computed, Jensen wake deficit model (wdm-JENS), which

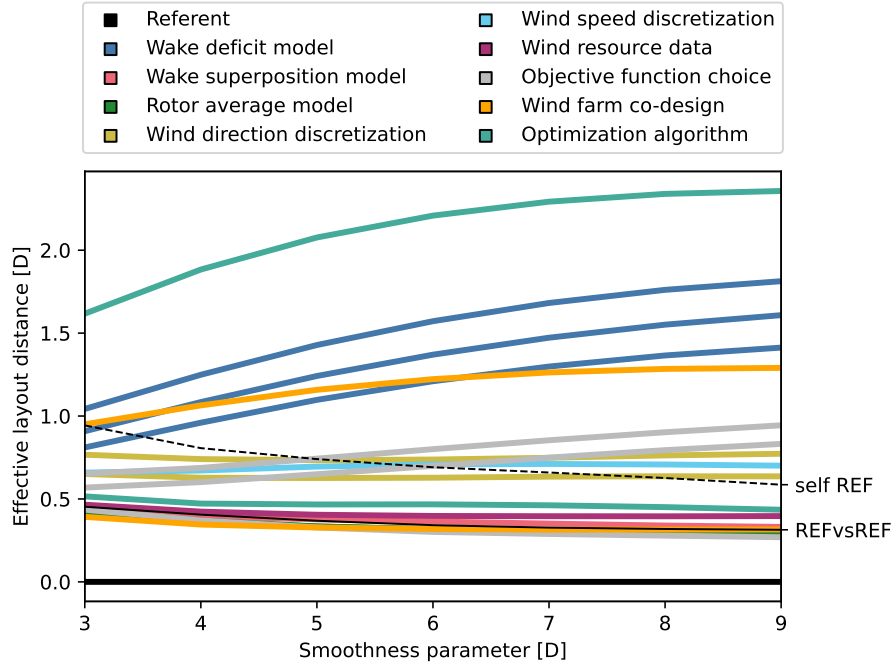


Figure A1. Impact of the choice of the smoothness parameter σ on the d_{eff} value calculated between the referent and the different implementation choices. The values of selfREF and REFvsREF are included as the dotted and the solid lines, respectively.

was previously identified as a high-impact choice, and the rotor center average model (ram-RCEN), which was identified as a low-impact choice. The scope is to verify whether the classification of these two alternative implementation choices as high and low impact, respectively, remains valid when a larger number of layouts is considered.

The results are presented in Fig. B1, where the d_{eff} values are reported for four different numbers of layouts, namely $N_{\text{runs}} = 5, 10, 15$, and 20.

It can be observed that selfREF remains approximately constant as N_{runs} increases. This indicates that the difference between one individual layout obtained with the referent implementation choice and a set of layouts obtained with the same implementation choice is already well captured for relatively small values of N_{runs} . Therefore, values of $N_{\text{runs}} \geq 5$ appear sufficient to obtain robust estimates of the intrinsic variability related to different runs for the referent. Then, REFvsREF decreases as N_{runs} increases, particularly for $N_{\text{runs}} \geq 10$, converging towards zero. This behavior is expected, as increasing the number of layouts in both sets reduces the influence of individual solutions, eliminating the differences between the two sets. This aspect further supports the robustness of the methodology, as the superimposed fields of the two sets converge towards being identical when a sufficient number of layouts is considered in each set.

The alternative implementation choice related to the wake deficit model (wdm-JENS), remains approximately constant for $N_{\text{runs}} \geq 10$. Therefore, its classification as a high-impact implementation choice is not affected by the number of layouts included in the comparison. Furthermore, the alternative rotor average model (ram-RCEN) follows the REFvsREF trend for

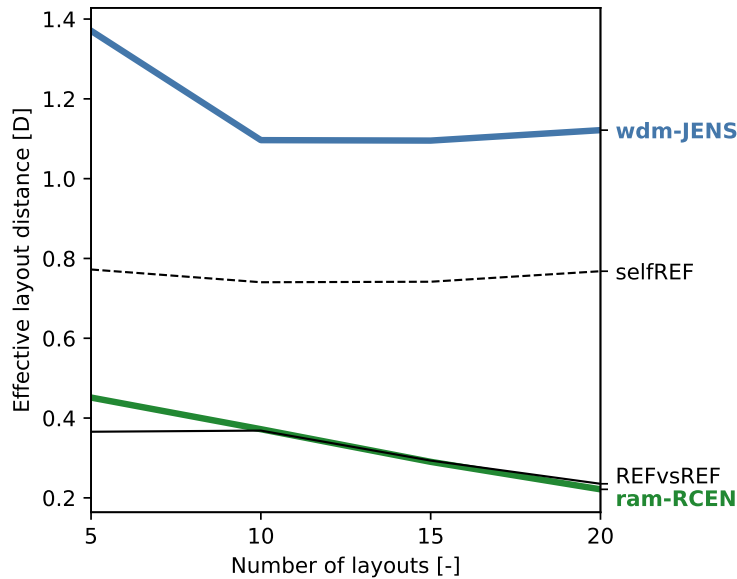


Figure B1. Impact of the number of layouts N_{runs} used for each comparison. The d_{eff} values between the referent and the two implementation choices wdmJENS and ramCEN are shown by the colored lines. The values of selfREF and REFvsREF are included as the dotted and the solid lines, respectively.

$N_{\text{runs}} \geq 10$, indicating that the effective layout distance between this implementation choice and the referent is comparable to the variability observed between two sets of layouts generated with the referent itself. This confirms its classification as a low-impact choice, independently of the selected number of layouts.

Overall, these results indicate that the classification of the two selected implementation choices would remain unchanged if more layouts were included in the comparison. The high- and low-impact choices considered here provide representative examples of how the results may vary with N_{runs} . Therefore, the conclusions obtained in this study using $N_{\text{runs}} = 10$ are expected to remain unchanged if a larger number of optimization runs were performed.

Author contributions. Matteo Baricchio: conceptualisation, methodology, software, validation, investigation, writing – original draft, visualisation. Jenna Iori: writing – review & editing, conceptualization, supervision. Pieter M.O. Gebraad: writing – review & editing, conceptualization, supervision. Jan-Willem van Wingerden: writing – review & editing, conceptualization, supervision, resources, funding acquisition.

Competing interests. At least one of the (co-)authors is an Associate editor of Wind Energy Science.



Financial support. This work has been supported by the SUDOCO project, which receives the funding from the European Union's Horizon Europe Programme under the grant No. 101122256.

630 *Acknowledgements.* The authors would like to thank Samuel Kainz for the valuable support with the use of DETECT software. The authors acknowledge the use of computational resources of the DelftBlue supercomputer, provided by Delft High Performance Computing Centre (<https://www.tudelft.nl/dhpc>), and that part of this work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative, using the Snellius supercomputer.



References

- Abramowitz, M. and Stegun, I. A.: Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, Dover Publications, New York, 1970.
- Azlan, F., Kurnia, J., Tan, B., and Ismadi, M.-Z.: Review on optimisation methods of wind farm array under three classical wind condition problems, *Renewable and Sustainable Energy Reviews*, 135, 110 047, <https://doi.org/10.1016/j.rser.2020.110047>, 2021.
- Baigarina, A., Kostas, K., and Cheng, P. W.: Impact of Wake Expansion Assumptions on Wind Farm Layout Optimization and Cabling Trade-offs, *Journal of Physics: Conference Series*, 3224, 032 097, <https://doi.org/10.1088/1742-6596/3224/3/032097>, 2026.
- Baricchio, M., Gebraad, P. M. O., and van Wingerden, J.-W.: Evaluating the potential of a wake steering co-design for wind farm layout optimization through a tailored genetic algorithm, *Wind Energy Science*, 9, 2113–2132, <https://doi.org/10.5194/wes-9-2113-2024>, 2024.
- Baricchio, M., Hoek, D. v. d., Dammann, T., Iori, J., Gebraad, P. M., and Wingerden, J.-W. v.: An approximation of the optimal combined helix and yaw control for wind farm co-design applications, *Journal of Physics: Conference Series*, 3224, 032 019, <https://doi.org/10.1088/1742-6596/3224/3/032019>, 2026a.
- Baricchio, M., Iori, J., Lokken, T. S., Gebraad, P. M., and van Wingerden, J.-W.: A novel metric to compare wind farm layouts and identify trends in the layout optimization problem, *Journal of Physics: Conference Series*, 3224, 032 020, <https://doi.org/10.1088/1742-6596/3224/3/032020>, 2026b.
- Bastankhah, M. and Porté-Agel, F.: A new analytical model for wind-turbine wakes, *Renewable Energy*, 70, 116–123, <https://doi.org/10.1016/j.renene.2014.01.002>, 2014.
- Dirik, D. G., Quick, J., Réthoré, P.-E., Murcia Leon, J. P., and Riva, R.: Revenue-Focused Wind Farm Control Co-Design for Future Electricity Markets Scenarios, *Journal of Physics: Conference Series*, 3016, 012 024, <https://doi.org/10.1088/1742-6596/3016/1/012024>, 2025.
- Frandsen, S.: Turbulence and turbulence-generated structural loading in wind turbine clusters, Ph.D. thesis, ISBN 87-550-3458-6, risø-R-1188(EN), 2007.
- Gao, X., Li, Y., Zhao, F., and Sun, H.: Comparisons of the accuracy of different wake models in wind farm layout optimization, *Energy Exploration & Exploitation*, 38, 1725–1741, <https://doi.org/10.1177/0144598720942852>, 2020.
- Gebraad, P. M. O., Kainz, S., Anand, A., Guillore, A., Quick, J., and Dirik, D. G.: Reference scenarios / D3.1, Tech. rep., SUDOCO, https://sudoco.eu/wp-content/uploads/2024/04/D3.1_Reference-Scenarios.pdf, 2024.
- Guo, N., Zhang, M., Li, B., and Cheng, Y.: Influence of atmospheric stability on wind farm layout optimization based on an improved Gaussian wake model, *Journal of Wind Engineering and Industrial Aerodynamics*, 211, 104 548, <https://doi.org/10.1016/j.jweia.2021.104548>, 2021.
- Göçmen, T., van der Laan, P., Réthoré, P.-E., Diaz, A. P., Larsen, G. C., and Ott, S.: Wind turbine wake models developed at the technical university of Denmark: A review, *Renewable and Sustainable Energy Reviews*, 60, 752–769, <https://doi.org/10.1016/j.rser.2016.01.113>, 2016.
- Italiano, N., Pettas, V., Göçmen, T., and Cutululis, N. A.: Towards holistic wind farm design optimization: an integrated multidisciplinary approach, *Journal of Physics: Conference Series*, 3224, 032 082, <https://doi.org/10.1088/1742-6596/3224/3/032082>, 2026.
- Jensen, N.: A note on wind generator interaction, no. 2411 in Risø-M, Risø National Laboratory, 1983.
- Jong, R. d., Doekemeijer, B. M., Kreeft, J. J., Verzijlbergh, R. A., and Wingerden, J. W. v.: Validation of real-weather LES using SCADA data from a large modern offshore wind farm, *Journal of Physics: Conference Series*, 3224, 032 003, <https://doi.org/10.1088/1742-6596/3224/3/032003>, 2026.



- 670 Kainz, S., Guilloiré, A., and Bottasso, C. L.: How do technological choices affect the economic and environmental performance of offshore wind farms?, *Journal of Physics: Conference Series*, 2767, 082 005, <https://doi.org/10.1088/1742-6596/2767/8/082005>, 2024.
- Mosetti, G., Poloni, C., and Diviacco, B.: Optimization of wind turbine positioning in large windfarms by means of a genetic algorithm, *Journal of Wind Engineering and Industrial Aerodynamics*, 51, 105–116, [https://doi.org/10.1016/0167-6105\(94\)90080-9](https://doi.org/10.1016/0167-6105(94)90080-9), 1994.
- NatLabRockies: FLORIS, <https://doi.org/10.5281/zenodo.20850486>, 2026.
- 675 Nygaard, N.: TurbOPark: A Turbulence Optimized Park model, Ørsted, <https://github.com/OrstedRD/TurbOPark>, 2022.
- O'Neill, N., Réthoré, P.-E., Mouradi, R.-S., Mathieu, A., and Quick, J.: Wind Farm Layout Optimization Accounting for Uncertainty in Model Selection, *Journal of Physics: Conference Series*, 3016, 012 054, <https://doi.org/10.1088/1742-6596/3016/1/012054>, 2025.
- Pedersen, M. M., Forsting, A. M., van der Laan, P., Riva, R., Romàn, L. A. A., Risco, J. C., Friis-Møller, M., Quick, J., Christiansen, J. P. S., Rodrigues, R. V., Olsen, B. T., and Réthoré, P.-E.: PyWake 2.5.0: An open-source wind farm simulation tool, <https://gitlab.windenergy.dtu.dk/TOPFARM/PyWake>, 2023.
- 680 Poole, D. J.: Characterization of Multimodality in Wind Farm Layout Optimization, *Energy Science & Engineering*, 14, 737–751, <https://doi.org/10.1002/ese3.70377>, 2026.
- Porté-Agel, F., Bastankhah, M., and Shamsoddin, S.: Wind-Turbine and Wind-Farm Flows: A Review, *Boundary-Layer Meteorology*, 174, 1–59, <https://doi.org/10.1007/s10546-019-00473-0>, 2020.
- 685 Quick, J., Rethore, P.-E., Mølgård Pedersen, M., Rodrigues, R. V., and Friis-Møller, M.: Stochastic gradient descent for wind farm optimization, *Wind Energy Science*, 8, 1235–1250, <https://doi.org/10.5194/wes-8-1235-2023>, 2023.
- Romero-Barrera, A. J., Casillas-Pérez, D., Cornejo-Bueno, L., Pérez-Aracil, J., Paricio-Garcia, A., Lopez-Carmona, M. A., Blanco-Sancho, J., Ghimire, S., Deo, R. C., Caamaño, A. J., and Salcedo-Sanz, S.: A comparative study of different kinematic wake models within meta-heuristics for efficient wind farm layout optimization, *Results in Engineering*, 26, 105 112, <https://doi.org/10.1016/j.rineng.2025.105112>, 2025.
- 690 Réthoré, P.-E., Fuglsang, P., Larsen, G. C., Buhl, T., Larsen, T. J., and Madsen, H. A.: TOPFARM: Multi-fidelity optimization of wind farms, *Wind Energy*, 17, 1797–1816, <https://doi.org/10.1002/we.1667>, 2014.
- Sanchez Perez Moreno, S.: A guideline for selecting MDAO workflows with an application in offshore wind energy, Phd thesis, Delft University of Technology, Delft, The Netherlands, <https://doi.org/10.4233/uuid:ea1b4101-0e55-4abe-9539-ae5d81cf9f65>, 2019.
- 695 Simpson, J., Loth, E., and Dykes, K.: Cost of Valued Energy for design of renewable energy systems, *Renewable Energy*, 153, 290–300, <https://doi.org/10.1016/j.renene.2020.01.131>, 2020.
- Souza de Alencar, M., Göçmen, T., and Cutululis, N. A.: Flexible cable routing framework for wind farm collection system optimization, *European Journal of Operational Research*, <https://doi.org/10.1016/j.ejor.2025.07.069>, 2025.
- Stanley, A. P. J. and Ning, A.: Massive simplification of the wind farm layout optimization problem, *Wind Energy Science*, 4, 663–676, <https://doi.org/10.5194/wes-4-663-2019>, 2019.
- 700 Stanley, A. P. J., Bay, C. J., and Fleming, P.: Enabling control co-design of the next generation of wind power plants, *Wind Energy Science*, 8, 1341–1350, <https://doi.org/10.5194/wes-8-1341-2023>, 2023.
- Thomas, J. J., Baker, N. F., Malisani, P., Quaeghebeur, E., Sanchez Perez-Moreno, S., Jasa, J., Bay, C., Tilli, F., Bieniek, D., Robinson, N., Stanley, A. P. J., Holt, W., and Ning, A.: A comparison of eight optimization methods applied to a wind farm layout optimization problem, *Wind Energy Science*, 8, 865–891, <https://doi.org/10.5194/wes-8-865-2023>, 2023.
- 705 Thomson, M. and Zaaijer, M.: Providing insight into what can be expected from Offshore Wind Farm Layout Optimisation, *Journal of Physics: Conference Series*, 2265, 042 043, <https://doi.org/10.1088/1742-6596/2265/4/042043>, 2022.



- Vortex FDC: Vortex Wind Map, Accessed via Youwind, <https://youwindrenewables.com>, accessed: 2024-11-01, 2024.
- Wiese, F., Bramstoft, R., Koduvere, H., Pizarro Alonso, A., Balyk, O., Kirkerud, J. G., Åsa Grytli Tveten, Bolkesjø, T. F., Münster, M., and
 710 Ravn, H.: Balmorel open source energy system model, *Energy Strategy Reviews*, 20, <https://doi.org/10.1016/j.esr.2018.01.003>, 2018.
- Yang, P. and Najafi, H.: The Effect of Using Different Wake Models on Wind Farm Layout Optimization: A Comparative Study, *Journal of
 Energy Resources Technology*, 144, 070 904, <https://doi.org/10.1115/1.4052775>, 2021.
- Zahle, F., Barlas, T., Lønbæk, K., Bortolotti, P., Zalkind, D., Wang, L., Labuschagne, C., Sethuraman, L., and Barter, G.: Definition of the IEA
 Wind 22-Megawatt Offshore Reference Wind Turbine, Tech. Rep. DTU Wind Energy Report E-0243, Department of Wind and Energy
 715 Systems, Technical University of Denmark, Roskilde, Denmark, <https://doi.org/10.11581/DTU.00000317>, 2024.
- Zhang, F., Wang, H., Zhang, Y., Miao, W., and Hou, C.: A Review of Mathematical Optimization Methods in Offshore Wind Energy Systems:
 Design, Layout, and Control, *Journal of Marine Science and Engineering*, 14, <https://doi.org/10.3390/jmse14111033>, 2026.
- Zong, H. and Porté-Agel, F.: A momentum-conserving wake superposition method for wind farm power prediction, *Journal of Fluid Me-
 chanics*, 889, A8, <https://doi.org/10.1017/jfm.2020.77>, 2020.