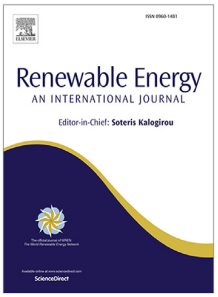
Journal Pre-proof



Topology-aware surrogate for future offshore wind farms using machine learning

Thuy-Hai Nguyen, Jean-François Toubeau, Emmanuel De Jaeger, François Vallée

PII: S0960-1481(25)01319-9
DOI: https://doi.org/10.1016/j.renene.2025.123657
Reference: RENE 123657

To appear in: *Renewable Energy*

Received date : 30 January 2025
Revised date : 30 May 2025
Accepted date : 2 June 2025

Please cite this article as: T.-H. Nguyen, J.-F. Toubeau, E.D. Jaeger et al., Topology-aware surrogate for future offshore wind farms using machine learning, *Renewable Energy* (2025), doi: https://doi.org/10.1016/j.renene.2025.123657.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2025 Published by Elsevier Ltd.

Revised Manuscript-Clear
Journal Pre-proof
Click here to access/download;Revised Manuscript-Clear;Renewable_Energy_R2_clear.pdf
Click here to view linked References

Topology-Aware Surrogate for Future Offshore Wind Farms Using Machine Learning

Thuy-Hai Nguyen[a,*], Jean-François Toubeau[a], Emmanuel De Jaeger[b], François Vallée[a]

[a] *Electrical Power Engineering Unit, University of Mons, 31 Boulevard Dolez, Mons, 7000, Belgium*
[b] *Department of Mechatronics, Electrical Energy and Dynamic Systems, University of Louvain, 2 Place du Levant, Louvain-la-Neuve, 1348, Belgium*

Abstract

As the role of offshore wind generation increases in modern power systems, the need for enhanced modelling techniques becomes critical. In problems involving iterative computations, simplified offshore wind farm models often disregard complex aerodynamic effects in order to reduce computational time. However, those effects have a significant impact on the produced electricity. To address this current modelling limitation, we develop a new topology-aware wind farm surrogate using supervised machine learning techniques. This surrogate model is trained to capture the complex relationship between free flow wind information and electrical output, irrespective of the wind farm layout. This allows bypassing the need to train a different model for each wind farm, while enabling the simulation of farms lacking historical data (such as those in construction phase). This is achieved by enriching the input space of the machine learning surrogate with novel geometric and physics-informed features, making it adaptable to any offshore wind farm configuration.

Keywords: Machine learning, Offshore wind, Surrogate, Topology-aware, Wake effects

*Corresponding author
Email address: thuy-hai.nguyen@umons.ac.be

Journal Pre-proof

Journal Pre-proof

1. Introduction

Offshore energy has the potential to deliver large amounts of low-carbon, renewable energy to fulfil a large part of future electrical needs. In 2023, the worldwide amount of installed offshore wind power was 75.2 GW [1]. The European Union has recently set targets for an installed capacity of at least 60 GW of offshore wind energy by 2030, and 300 GW by 2050. However, the output power of offshore wind farms is weather-dependent, and thus highly fluctuating, and hard to predict. This leads to growing concerns regarding the reliability of the electricity supply for future renewable-dominated power systems. In this paper, we aim to enhance the quantification of these issues by improving the modelling of offshore wind farm operations.

Within a wind farm, the wind speeds seen by different turbines vary due to complex aerodynamic phenomena. Wake effects cause a reduction of wind speed due to upwind turbines extracting kinetic energy from the wind, and turbulence influencing the recovery of wind speed. These effects are particularly significant in offshore conditions, thus having a considerable impact on the resulting wind farm generated energy. Modelling the exact output of offshore wind farms in problems involving iterative computations is often performed using very simple wind farm models. Such problems applied to power systems include adequacy computations that aim at assessing the reliability of a power system, and involve simulating millions of possible states for the power system. Therefore, for iterative power systems computations, wake losses are either ignored [2], or roughly approximated [3, 4, 5], notably with the Jensen wake model [6].

Accurately assessing the power production of a wind farm, i.e. accounting for aerodynamic effects and wind turbines interacting with each other, is not straightforward. However, given the planned growth of offshore wind farms, it is important to understand and accurately simulate the variability of the output power with a tractable computation time in order to ensure their reliable integration into power systems. Wind farms simulation models can accurately capture complex wake effects, but their high computational burden makes them impractical. However, a database can be generated from wind farm simulations, then be used to train fast models based on Machine Learning (ML) techniques. In this paper, supervised ML techniques will thus be used to build a fast, accurate, and topology-aware surrogate model of offshore wind farms.

Journal Pre-proof

Most current works [7, 8, 9, 10] using ML to represent wind farms are building a model *only applicable for the wind farm on which they have been trained*, i.e., for a fixed layout and wind turbine characteristics. This means that the model has to be completely retrained for any modification of the wind farm configuration, i.e., adding/removing turbines or changing the turbine type in case of re-powering of older turbines. Consequently, the current ML models cannot be applied to various wind farms, such that a different model has to be trained for each wind farm topology. Considering the diversity in layout shape, turbine spacing, turbine characteristics, ... of the future offshore wind landscape, current methods are not computationally efficient. Moreover, the topology of future wind farms, which are not yet built, remains unknown and this uncertainty must be taken into account when studying them, which can be achieved by considering different possible configurations (leading to different power outputs) of those farms. Finally, none of the cited works allows the possibility to consider turbine failures. Indeed, in case of maintenance or repair, an idle turbine does not produce electricity, thus does not extract energy from the wind. Its situation affects all downwind turbines, and the wind energy is actually redistributed among neighbouring turbines in a complex way. This impacts the wind farm production, especially with turbines that have a large rotor diameter.

A few papers have focused on creating generalized ML models, to ensure that the training process should be carried out only once. In [11], two neural networks (respectively capturing wake and turbulence field) are trained with a database consisting of simulations performed on a standalone wind turbine. While showing strong capabilities to capture the complex spatial relationship between inflow conditions and the wake fields, the developed model exhibits a high computation time for large wind farms. Indeed, power computations for a wind farm requires the propagation of the wake field, which involves running the neural network many consecutive times. This can lead to error accumulation as the power prediction on a downstream turbine depends on the predictions (and accuracy) made on the upstream turbines. In reference [12], a dataset is built upon medium-fidelity aerodynamic simulations (based on Reynolds-averaged Navier Stokes equations) and the ML model uses atmospheric and physics-informed features to generalize to different layout configurations. However, the proposed model is limited to very specific wind conditions. Authors in [13] train physics-induced graph neural networks, able to generalize to unseen wind farm configurations. However,

Journal Pre-proof

the use of graph neural networks involves a high computation time that increases with the number of turbines. In this way, the reported time is 0.1 s for a 40-turbine wind farm, which exceeds wind farm simulation tools based on advanced wake models, perhaps making them quite cumbersome for iterative computations. Besides, the model in [13] cannot handle heterogeneous turbine types.

In this context, this paper presents a novel approach to develop a surrogate model for offshore wind farms, built on wind farm simulations. The model accounts for heterogeneous turbine types and turbine failures and, most importantly, is designed to generalize to unseen wind farm configurations. This means it can be applied to any offshore wind farm worldwide, with the training process performed only once. Furthermore, the model achieves low computational times for power assessments, making it highly efficient and broadly applicable. Specifically, this paper has three main contributions.

Firstly, a generic, topology-aware ML model is trained to accurately capture complex aerodynamic effects arising within offshore wind farms. The model relies on an augmented set of features, including atmospheric, geometric, physics-informed, and turbine-specific characteristics, which makes it able to generalize to any (future) farm configuration. By externalizing the complexity from the model architecture to the input features set, the proposed model has a very low computation time in both training and inference stages, such that the ML model is highly valuable in problems involving iterative computations (see methodology in Fig. 1).

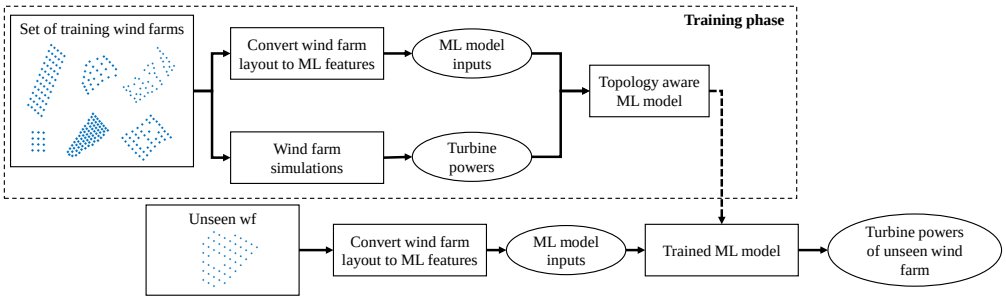


Figure 1: Diagram of the proposed methodology

Secondly, a database of is built to form a significant training set for the ML surrogate. In the absence of real-life data, offshore wind farm simulations

Journal Pre-proof

are leveraged to constitute the training database. The latter encompasses enough diversity in wind conditions and wind farm configurations for enabling the ML model to generalize to unseen data. Situations with turbine failures and the subsequent redistribution of wake are also encompassed within the database.

Thirdly, the generalization capabilities of the model are leveraged to account for uncertainties in the layout and turbine choice of future offshore wind farms. A sensitivity analysis on yearly energy production with regard to farm density and layout, as well as turbine characteristics, can be easily carried out with the topology-aware surrogate. This is further illustrated in our case study by integrating the model into a long-term adequacy study aimed at assessing the reliability indices of a future power system with a high share of offshore generation.

The remainder of this paper is organized as follows. Section 2 explains the creation of the training database using wind farm simulations. In section 3, the generic wind farm surrogate is presented: the different input features are explained and the performances of the proposed model are evaluated on a real wind farm. Benchmark comparisons with existing models are carried out, and the scalability to large wind farms is also studied. In section 4, a case study of adequacy studies of the Belgian power system using the Monte-Carlo methodology is performed. Finally, main findings and perspectives are summarized in the last section.

2. Wind farm simulations: generation of a representative synthetic database

As wind farm simulations (even with low or medium-fidelity models) cannot be directly embedded in complex power systems problems, there is a crucial need for creating a representative synthetic database on which faster ML models can be trained.

2.1. Wind farm simulations

To build the training database, Computational Fluid Dynamics (CFD) would provide very accurate wind farm simulations, but it is very complex and computationally too demanding. Therefore, as a good trade-off between high accuracy and tractability, the choice was made to use lower fidelity wake models (or engineering models). It should be noted that the use of

Journal Pre-proof

wake models to build the training database is not the main focus of this work. Indeed, the most accurate way to capture aerodynamic effects would be to rely on measurements data, which encompass not only wake effects, but also other complex phenomena such as blockage. Unfortunately, those data are not currently available in the literature, even though they exist: vast quantities of information are collected by turbine manufacturers, operators and utility companies [14]. In this paper, the use of wake models is not critical, as it only serves to replace measurements data that are not yet accessible. Those data could be noisy and a careful pre-processing [15] (e.g., filter down-regulation, startup, or shutdown events) is needed before using them for training machine learning models.

Several commercial (WAsP, windPRO) and open-source softwares (FLORIS [16], PyWake [17], FOXES [18]) implement a common set of wake models. However, authors in [19] found a good agreement across most of the model implementations. In this work, simulations are run with FLORIS, though other tools could have been employed without significantly altering the results. FLORIS (FLOw Redirection and Induction in Steady state) [16] is developed by the National Renewable Energy Laboratory and it is widely used in the literature [20, 21]. This open-source code provides a modelling tool of the steady-state wake characteristics in a wind farm. Floris offers several wake models: in this study, gaussian models developed in [22] and based on the implementation by Niayifar and Porté-Agel are chosen for wake deflection and velocity deficit, as they have been proven to be in good agreement with both numerical and experimental results. Indeed, the Gaussian shape of the hub height velocity deficit in the far wake can clearly be seen in wind tunnel measurements [23, 24], numerical simulations [25], and data of operating wind farms [26]. Furthermore, the wake model is complemented with a turbulence model to capture the turbulence field, which is essential in the context of interactions among multiple turbines. Here, the Crespo-Hernandez approach [27] is selected for modelling the added turbulence due to turbine operation. Finally, depending on the wind direction, the turbines inside wind farms are often exposed to multiple wakes from several upstream wind turbines. These overlapping wakes are combined using freestream linear superposition. The number of grid points (where the wind conditions are calculated) on each turbine rotor was set to 9. Blockage effects are not modelled as they do not to have a critical impact on wind energy utilization. Indeed, global blockage has a much lower impact on the power production

Journal Pre-proof

than other wind farm flow features like inner-farm wakes. [28]. For the six real-life offshore wind farms simulated for the training set (see next section), blockage losses (computed with PyWake) do not exceed 0.5% (1% for Northwind only).

We have validated our parametrized Floris-based model by comparing its outputs with actual measurements from two offshore wind farms. As more thoroughly detailed in [10], this comparison spanned both farm-level and turbine-level outputs across various wind speeds and directions, and demonstrated that our simulation model accurately captures the width and depth of power losses due to wake effects. Moreover, as Floris offers a great flexibility in the choice of wake models, one could choose another combination of velocity deficit and superposition models to build another database. Since our model is topology-aware, the training process should be carried out again only once. While the developed surrogate will be linked to the training set and thus to the choice of wake models, our methodology can be considered as model-agnostic.

2.2. Database generation

Our Floris-based simulation model is then used on six real-life offshore wind farms for generating the training database (they can be seen in Fig. 2):

- Alpha Ventus: 12 turbines, 4 km^2, 60 MW (15 MW/km^2)
- Belwind: 56 turbines, 17 km^2, 171 MW (10 MW/km^2)
- Mermaid: 28 turbines, 17 km^2, 235 MW (13.8 MW/km^2)
- Norther: 44 turbines, 44 km^2, 370 MW (8.4 MW/km^2)
- Northwind: 72 turbines, 15 km^2, 216 MW (14.4 MW/km^2)
- Rentel: 42 turbines, 23 km^2, 309 MW (13.4 MW/km^2)

They are all located in the North Sea but they offer variable characteristics in terms of layout and turbine technology. Their layouts show different wind turbine density, shape and number of turbines. The wind turbines have different rotor diameter, hub height and power curve shapes (see Table 1). For each farm, the layout (wind turbine coordinates) is publicly available [29, 30], and the wind turbine characteristics are extracted from manufacturers datasheets. Two wind farms (Belwind and Alpha Ventus) have heterogenous turbine types. The purpose of using only 6 layouts for the training set is to allow to build a ML surrogate showing topology-aware abilities even with a limited amount of wind farms data.

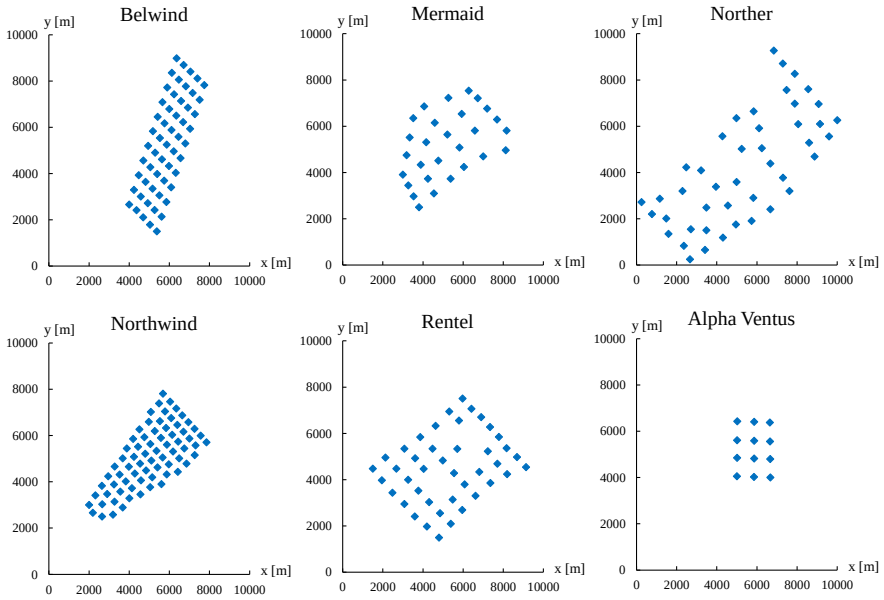


Figure 2: Wind farms used for training the ML model

	Type	Rotor diameter	Hub height	Rated power	Rated wind speed
Alpha Ventus	Adwen	116 m	90 m	5 MW	12.5 m/s
	Senvion	126 m	94 m	5 MW	14 m/s
Belwind	Vestas	90 m	72 m	3 MW	15 m/s
	Heliade	150 m	100 m	6 MW	13 m/s
Mermaid	Siemens	167 m	109 m	8.4 MW	14 m/s
Norther	Vestas	164 m	105 m	8.4 MW	13 m/s
Northwind	Vestas	112 m	71 m	3 MW	13 m/s
Rentel	Siemens	154 m	105 m	7.35 MW	13 m/s

Table 1: Wind turbine characteristics

Regarding the list of wind speeds, wind directions and ambient turbulence, the ERA5 dataset provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) [31] was used. ERA5 is based on reanalysis to generate atmospheric time series covering a large period (from 1959 to 2024), with a spatial resolution of 31 km and an hourly temporal resolution.

Journal Pre-proof

Reanalysis combines model data with observations from across the world into a globally complete and consistent dataset using the laws of physics. It provides a comprehensive description of the observed climate as it has evolved during recent decades, on 3D grids at sub-daily intervals. Since ambient turbulence intensities are not directly given by the ERA5 dataset, the IEC standard 61400-3 [32] formulation is used to get realistic turbulence values.

For each wind farm, one year (2019) of hourly wind along with extreme scenarios (very high wind speeds) at the location of the offshore wind farms were used as atmospheric inputs. Hence, a total of 52,746 (6*8791) simulations were run with Floris. The resulting database consists not only in the aggregated power output of the entire wind farm, but also the individual wind turbine outputs, **taking into account potential turbine failures in the farm**. Those failures were randomly sampled from a probabilistic distribution based on offshore turbines reliability data.

3. Wind farm surrogate using machine learning

To train the ML surrogate, the database built in section 2.2 is used. Each wind turbine of every wind farm simulation is considered as a sample. Therefore, without pre-processing, each data sample represents an hourly value of wind speed, wind direction and ambient turbulence, and the corresponding effective wind speed (i.e., the wind speed seen right before the rotor, which can be directly used to get the wind turbine steady-state power). We predict the output of a single wind turbine because an evaluation at the turbine level gives a better insight of the ML models performance since turbine errors may compensate when powers are aggregated into total power at the farm level. Moreover, this allows to consider individual wind turbine outages when the model is used in adequacy studies.

The main challenge is to ensure that the trained ML model is layout-agnostic, i.e., the model should be able to make power assessments for a farm unseen in the training database. An important contribution of this work is therefore to enrich the input space with new features, including atmospheric, geometric, physics-informed, and turbine-specific information, which enable the ML model to generalize to any farm configuration. Moreover, while most ML models in the literature aim at predicting turbine power production, our ML surrogate directly outputs the waked wind speed u_{WT_i}, i.e., the wind speed seen at the rotor of the i^{th} turbine WT_i. The wind speed u_{WT_i},

Journal Pre-proof

as opposed to the free-flow speed, takes into account the influence of all upstream turbines on the free-flow wind speed (i.e., velocity deficit from the wake, but also wake recovery due to turbulence). Indeed, this allows our ML surrogate to generalize to turbine with different characteristics (e.g., different rotor diameter). To obtain the wind turbine power output, the waked wind speed is fed to the turbine power curve, then the wind farm total power is computed by summing individual turbine outputs. The process is detailed in Fig. 3. Practically, each sample consists of 16 input features, which are detailed hereafter, while the direct output of the ML model is the wind speed seen at the turbine rotor.

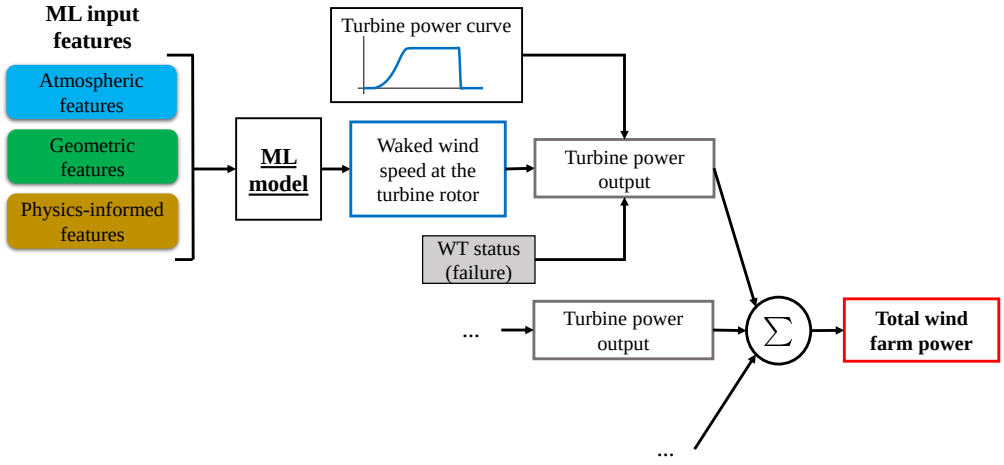


Figure 3: Diagram of the methodology used to predict wind farm power with the topology-aware ML model

3.1. Features for topology-aware wind farm model

3.1.1. Atmospheric features

Wind speed in front of the rotor directly determines the output generated by the turbine. However, this information is not readily available, so the free-flow wind speed measured ahead of the wind farm is used. When taking wake effects into account, wind direction also has a crucial impact. However, wind direction is not directly used as a feature, but rather implicitly considered when computing the geometric and physics-informed features, as explained hereafter. Ambient turbulence intensity is also fed to the model.

Journal Pre-proof

3.1.2. Geometric features

For each sample (corresponding to a specific time step for a wind turbine within a farm), the computation of geometric features allows to convert information regarding wind direction and wind farm layout into usable inputs for the ML model. Based on the knowledge on how wake effects propagate (wake expansion will be addressed in section 3.1.3) and affect downstream turbines, the following features have been computed for every sample:

- Number of upstream blocking wind turbines

  This feature indicates the number of turbine located upwind, i.e., whose swept area (or part of) coincides with the rotor of the considered wind turbine, affecting non-linearly its wind access and performance. No blocking turbine means that the considered turbine is unwaked, and produces the maximum power for the free flow wind speed. However, more blocking turbines means less available energy in the wind seen by the rotor.

- Mean distance of blocking wind turbines

  The distances between the considered wind turbine and all upstream blocking turbines are averaged. Distances between turbines are computed perpendicularly to the rotor areas, in the direction of the flow. A high number may indicate a lower turbine density for the wind farm (turbines installed at a larger distance from one another), thus allowing for more wake recovery.

Wind farm measurements have shown that the closest upwind turbine aligned with the considered rotor has the most influence on power deficit [33]. Therefore, we identify the closest blocking turbine $WT_{b,1}$ of turbine WT_i in terms of streamwise direction (i.e., direction of mean flow).

- Stream-wise distance of first blocking wind turbine

  This is the distance between the considered wind turbine WT_i and the closest upstream blocking turbine $WT_{b,1}$, in the streamwise direction.

- Lateral distance to first blocking turbine

  This distance is the shift between the considered wind turbine WT_i and the closest upstream blocking turbine WT_b, in the spanwise direction (i.e. direction perpendicular to the mean flow). Indeed, the velocity

Journal Pre-proof

deficit will be maximum if the blocking rotor is completely shadowing the considered turbine. This distance is computed between rotor centrelines.

- Rotor diameter of first blocking turbine

  A turbine with a larger rotor diameter generally extracts more power from the incoming wind, as turbine power production is proportional to the square of the rotor radius. Therefore, for the same streamwise distance, a larger turbine will lead to a increased velocity deficit in the wake, thus reducing the waked wind speed seen at the rotor of the considered turbine WT_i.

There may be multiple upstream turbines that block the same or additional portions of the swept area of a given turbine. Moreover, as it was stated before, blocking turbines that are closer have more influence on the waked wind speed at the rotor of turbine WT_i. However, if those turbines are not completely aligned with the considered rotor, their influence decreases. The combination of those two aspects is captured by the blocking ratio and the blocking distance described below.

- Blocking ratio

  This is the fraction of the swept area of turbine i that is blocked by the swept area of any wind turbine upstream [34]. This blocking ratio BR_i varies between 0 (not blocked) to 1 (totally blocked) and is computed as:

$$BR_i = \frac{1}{A_i} \int_{(x,y)\in A} \chi \, dx \, dy \tag{1}$$

  where

  A_i is the swept area of the rotor of turbine WT_i

  $\chi = \chi(x, y)$ is a function defined for each point (x, y) on the turbine rotor disk.

  We have $\chi = \begin{cases} 1, \text{ if the point } (x, y) \text{ is blocked by any turbine.} \\ 0, \text{ otherwise.} \end{cases}$

- Blocking distance

Journal Pre-proof

It represents the weighted average distance to the upstream blocking turbines. The blocking distance BD_i of turbine WT_i is defined as:

$$BD_i = \frac{1}{A_i} \int_{(x,y)\in A_i} [\chi \ \min(L, L_\infty) + (1-\chi) \ L_\infty] \, dx \, dy \tag{2}$$

where

L is the distance between the grid point and any upwind blocking turbine

$L_\infty = 20 * D$ (for grid points not blocked by any upwind turbine)

The blocking distance can be seen as the average distance to the upstream blocking turbines, weighted by the fraction of the area obstructed by each upstream blocking turbine. The term L_∞ is an arbitrary large but finite length ensuring that the integral remains finite. In this study, L_∞ is set to a distance equal to 20 rotor diameters since previous studies have shown that the velocity profile in a turbine wake almost entirely recovers its original shape at 20 rotor diameters downstream of the turbine [35].

For the computation of these features, the swept area of each turbine is divided into a large number of discrete areas N_A, and the integrals in Eqs.(1) and (2) are evaluated numerically to give the geometric measures. The quantities BR_i and BD_i are computed at the geometric centre of each discrete area with the following equations:

$$BR_i = \frac{1}{N_A} \sum_{i_A=1}^{N_A} \chi_{i_A} \tag{3}$$

$$BD_i = \frac{1}{N_A} \sum_{i_A=1}^{N_A} [\chi_{i_A} \ \min(L, L_\infty) + (1-\chi_{i_A}) \ L_\infty] \tag{4}$$

where χ_{i_A}=1 if the centre of the area i_A is blocked by any upwind turbine, and 0 otherwise.

Journal Pre-proof

3.1.3. Physics-informed features

Complex ML models (such as deep neural networks, or random forests with numerous deep trees) can be plagued by a black-box nature, which makes them difficult to interpret and prone to produce physically impossible results. A recent field of research that has focused on solving this problem is physics-informed machine learning [36]. Physics-informed machine learning is characterized by the exploitation of the scientific knowledge or the physics laws to guide the optimization, architecture design, and implementation of ML models. In particular, physics-informed feature engineering refers to leveraging the insights of prior physical knowledge to select, transform, and synthesize the most informative input variables from raw data for the ML model [37].

In this paper, we want to improve the model performance by efficiently capturing that geometric features simply describe the layout of the wind farm, and it uses turbine rotor diameters to compute the swept area of the upstream turbine rotor. This equates to an assumption that the wake behind a turbine has a constant diameter equal to the rotor diameter of this turbine. However, it is known that the wake expands behind a turbine, meaning that the diameter of the wake increases with the downstream distance [38]. Indeed, when using only geometric features, the ML model may infer that some turbines are exposed to free-flow wind (if the number of blocking turbines found using rotor diameters is equal to 0), while they are in fact slightly in the wake of another turbine. As depicted in Fig. 4, this could happen if the turbines are not directly aligned such that the downstream turbine is not in the wake cylinder whose cross section is computed using the upstream turbine rotor diameter, but within the cylinder of a an expanded wake.

Determining the expanded wake diameter at the downstream turbine rotor has physical significance, aligning the ML model with aerodynamic principles. However, this calculation must be fast and simple to preserve the computational efficiency of the ML model. This is why we use the very simple Jensen wake model [39] to compute the wake diameter D_w. The Jensen model assumes a top-hat shape for the normalized velocity deficit and was derived using only mass conservation. The Jensen wake diameter is computed as follows:

$$D_w = D + 2 * k_w * x_w \tag{5}$$

where k_w is the Jensen wake spreading parameter, D the wind turbine diameter and x_w the distance downstream behind the turbine. In offshore

Journal Pre-proof

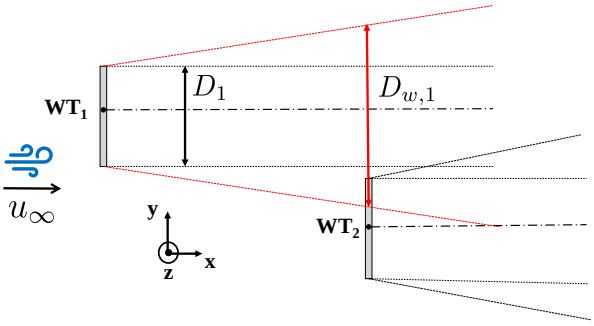


Figure 4: Illustration of geometric features in a simple case with 2 turbines. Plan view with where rotor diameters are represented by grey rectangles and their swept area by dotted lines. The wake expansion is shown by the red dashed lines.

conditions, k_w is usually set to 0.05 [40]. Therefore, additional values for the number of blocking turbines, mean blocking distance, distance/shift/rotor diameter of first blocking turbine, blocking ratio and blocking distance are computed using a wake diameter D_w instead of turbine rotors D. These new inputs are called physics-informed features.

3.2. Supervised ML algorithms

The relationship between the effective wind speed of each turbine and the proposed input features is highly non-linear, which motivates the use of the four following supervised ML algorithms:

- Decision tree
- Random forest
- Gradient Boosting Regression Tree (GBRT)
- Neural network (more specifically, Multi-layer Perceptron, NN-MLP)

3.3. Performance of the ML Surrogate

The synthetic database of section 2.2 was used to train the topology-aware ML model. To find the best (combinations of) hyperparameters for the supervised ML algorithms, a validation dataset of wind farm simulations is set up. The latter consists in hourly simulations performed on the Lillgrund offshore wind farm, i.e., a new wind farm unseen in the training database. It uses Floris and one year (2020, i.e., another year than for the weather data seen during training) of ERA5 wind data. Lillgrund is located near the

Journal Pre-proof

coast of Sweden, and the layout of the wind farm is presented in Fig. 5(a). This wind farm has a total rated capacity of 110 MW and consists of 48 pitch-controlled, variable speed wind turbines with a rotor diameter of 93 m and a hub height of 65 m. The layout is quite tight, with a mean turbine spacing of 3.3 rotor diameters. The hyperparameters of each ML model were calibrated with a thorough grid search performed on the validation set, and the obtained optimal architectures are given in Table 2.

Decision tree	Maximum depth	20
Random forest	Number of trees	50
	Maximum depth	25
Gradient boosting regression trees	Number of trees	50
	Maximum depth	15
	Learning rate	0.1
Multi-layer perceptron	Number of hidden layers	2
	Neurons in hidden layers	50
	Activation function	ReLU

Table 2: Values of hyperparameters calibrated on the validation set

Moreover, to quantify the performances of the ML surrogate and assess its ability to generalize to unseen data, the model is evaluated on a test set, which is composed of a new wind farm (not used in training and validation stages). The latter consists in hourly simulations performed on the large offshore wind farm Horns Rev (widely studied in the literature), using Floris and one year (2021) of ERA5 wind data. Horns Rev is located in the eastern North Sea, near the coast of Denmark. It consists of 80 wind turbines (Vestas V-80), whose rated power is 2 MW, hub height is 70 m above sea level and rotor diameter is 80 m. The wind farm layout (Fig. 5(b)) has a rhomboid shape with a minimum spacing of seven rotor diameters between two consecutive turbines. The total area of the wind farm is about 20 km^2 and the total installed capacity is 160 MW, meaning a power density of 8 MW/km^2. Similar to the training database, random turbines failures are also added to the test set. The performance of each ML model is given in Table 3. The Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE) are computed and given in MW but other metrics are given in percentage of the wind farm total installed capacity (160 MW for Horns Rev), i.e., Root Mean Squared Percentage Error (RMSPE) and Mean Absolute Percentage Error

Journal Pre-proof

(MAPE). The inference time corresponds to the total time for obtaining a yearly output of 8760 samples. The chosen algorithm should reflect the best compromise between accuracy, inference time and model complexity.

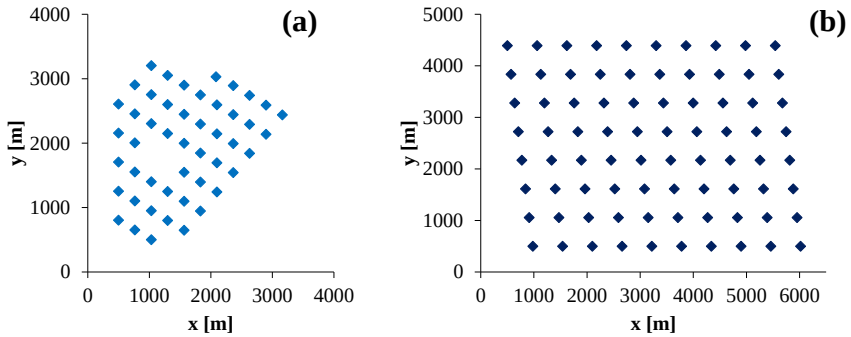


Figure 5: Layout of the (a) Lillgrund and (b) Horns Rev offshore wind farms

	Tree	RF	GBRT	NN-MLP
RMSE [MW]	2.47	2.23	1.93	1.57
RMSPE [%]	1.54	1.40	1.20	0.98
MAE [MW]	1.28	1.07	1.05	0.84
MAPE [%]	0.80	0.67	0.66	0.53
Inference time [s]	0.31	2.09	0.32	1.77

Table 3: Performance of machine learning models on the test set

The four ML models show strong performance regarding both accuracy and inference time. The decision tree has a very low prediction time, but has the highest RMSE (decision trees are prone to overfitting). Random forests exhibit lower errors, but at the cost of a higher inference time. The GBRT algorithm seems to offer a good compromise between accuracy (low errors, with RMSPE around 1% of wind farm total capacity) and computation time. The MLP is slower, but shows lower errors than random forests and GBRT, with both RMSPE and MAPE under 1% of the wind farm rated capacity. Neural networks (MLP) are thus the supervised algorithm chosen for rest of this work. Throughout the remainder of this paper, the term "ML surrogate" will refer to the model trained using MLP.

It should be noted that the time needed to calculate geometric and physics-informed features should be added to the inference time reported

Journal Pre-proof

in Table 3. For Horns Rev, the computation time is around 10.3 s, which brings the total inference time between 10.6 s and 12.39 s. This value is much lower than the 332 s needed to carry out 8760 simulations (i.e., one year) of Horns Rev with Floris. This means that the ML model is more than 25 times faster than wind farm simulations, making it ideally suited for computations where wind farm total power has to be simulated for thousands or even millions iterative steps.

Moreover, if our surrogate was to be used in power systems computations involving gradient-based methods, it should be differentiable. For decision trees, this could be accomplished by using soft decision trees or deep neural decision forests. For neural networks, it involves using smooth activation functions or ReLU (if the derivative is set to 0 at the origin [41]). For example, the gradient of total wind farm power with regard to the wind turbines positions was computed for our surrogate in [42].

3.3.1. Benchmark study with existing models

We compare the performance of our topology-aware surrogate with other well-established models based on ML and exhibiting capabilities to generalize to unseen wind farm configurations. This benchmark study is detailed in Table 4, where the performance is compared in terms of MAPE. We can see that our surrogate performs well with regard to the other models presented in the literature, especially when applied on an unseen wind farm (testing). However, it should be noted that the comparison is not straightforward. Indeed, the training database was not generated in the same way for all compared references: [12] is based on CFD simulations, [43] on measurements, while [13] and [44] use wake models, as does this work. Therefore, performance metrics reported in [43] also encompass measurements errors, and engineering model errors are included in [13, 44]. Moreover, the number of data in the testing set is different in all references. Regarding inference time, information on computational time was not provided in the other works, except in [13]. In the latter, making one power prediction for a wind farm of 40 turbines took approximately 0.1 s, which is slower than our surrogate (12.39 s for 8760 power predictions of a 80-turbines farm, i.e., 0.0014 s per prediction). Finally, references [44] and [43] acknowledge that while their models can generalize to unseen turbine positions, they cannot generalize to different turbine technologies. It is important to keep in mind that considering the discrepancies in methodologies and datasets, no direct conclusions can be made from this comparison: compared values are merely given for information purposes only.

Journal Pre-proof

Reference	Wind farm	MAPE
[12]	Horns Rev (training)	2.27%
	Horns Rev (testing)	4.21%
	Lillgrund (testing)	6.58%
[13]	Random farms (training)	1.58%
	Random farms (testing)	2.91%
[44]	Horns Rev (testing)	3.15%
[43]	Lillgrund (training)	3.5%
This paper	Horns Rev (testing)	0.53%

Table 4: Benchmark comparison with recent models found in the literature for wind farm power prediction using ML techniques

3.3.2. Scalability and generalization on irregular layouts

While the configuration of the Horns Rev wind farm was not seen during the training phase, both in terms of turbine positioning, turbine spacing and turbine technology, it would also be interesting to assess how our surrogate performs on a very irregular layout. Indeed, Horns Rev exhibits a rather regular layout, where turbines are aligned on equally distant rows. Moreover, it could also be relevant to assess the scalability of our developed model to larger wind farms. Indeed, as the use of offshore wind energy expands worldwide, the wind energy industry is considering building larger clusters of turbines [45]. Simulations were run with Floris for one year (2021) of ERA5 wind data, for a large wind farm of 160 turbines (same technology than the ones installed in Horns Rev) with an irregular layout, depicted in Fig. 6.

When making power predictions for the total farm output, the ML surrogate exhibits a RMSE of 8.51 MW and a MAE of 5.23 MW. Since the total installed capacity is 320 MW, those errors correspond to respectively 2.66% and 1.63% of the farm rated power. In terms of annual energy production (AEP), the AEP computed with Floris is 837.84 GWh, while it is 872.4 GWh with the surrogate: the relative error is 4.13%. Using day-ahead electricity prices for Belgium in 2021 (available on the ENTSO-E transparency platform [46]), we compute a relative difference in revenues of 4.15%. Therefore, even for large wind farms with an irregular layout (unseen in the training database), error metrics remain low.

Moreover, to ensure that errors in individual turbine power productions

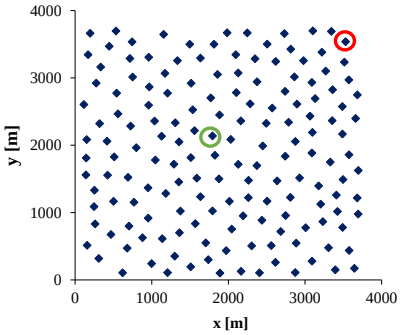


Figure 6: Large wind farm with irregular layout (centre turbine in green, edge turbine in red)

do not cancel out when summing to obtain the total farm power, we also investigate the ML surrogate performance on turbine power output. To that end, the normalized power with regard to wind direction is investigated for two turbines: one in the centre (highlighted in green in Fig. 6), and one at the edge of the farm boundaries (highlighted in red). It can be seen in Fig. 7 that the ML surrogate is able to accurately capture the power variations due to wake effects. For the turbine located at the centre of the farm (Fig. 7(a)), all directions are associated with wake losses, which clearly highlight the importance of accounting for such power losses in that case.

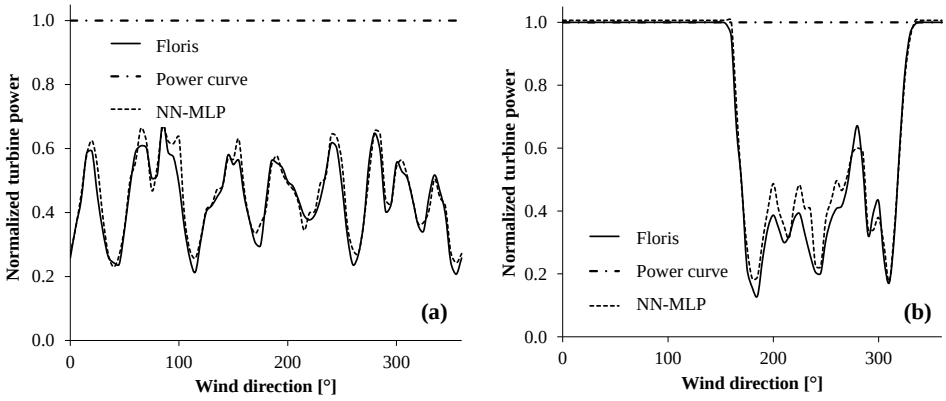


Figure 7: Comparison of normalized turbine power production at a large wind farm with irregular layout, for a wide sector of wind directions and a wind speed of 8 m/s; (a) for a turbine located at the centre, (b) for a turbine on the edge

Journal Pre-proof

4. Application of the topology-aware ML surrogate to adequacy studies: a case study on Belgium

One important component of the reliability of electricity systems is adequacy, i.e., ensuring that the total available generation capacity is sufficient to meet demand at all times [47]. Adequacy studies aim to compute reliability indices for estimating the systemic risk of electricity shortage, and sequential Monte-Carlo simulations stand as one of the most accurate methods for performing adequacy computations [48]. The sequential approach allows to properly represent the time-varying patterns in hourly generation and load, which makes it perfectly suited for the study of intermittent generating sources such as offshore wind generation [49]. The resulting iterative process is associated with a computational burden, which may limit the use of advanced models to represent complex renewable energy sources.

The adequacy indices computed in this work are the **Loss Of Load Expectation (LOLE)**, expressed in a number of hours per year [h/y] and the **Loss Of Energy Expectation (LOEE)**, in energy per year [GWh/y] [47, 50]. The LOLE is defined as the expected number of hours in the year during which the electricity consumption exceeds the production, whereas the LOEE computes the energy not served.

The test case presented in this section is an adequacy study of a simplified version of the Belgian power system foreseen in 2030. The interested reader can find more details on data relevant for this case in [3]. As Belgium is expected to build new offshore wind farms by 2030, the topology-aware surrogate is particularly appropriate for such a long-term adequacy study. Indeed, there is a high uncertainty on how wind farm developers will optimize their installation and it is thus important to consider multiple scenarios for this future offshore zone. Since our wind farm surrogate is topology-aware, the training process is carried out only once, then the model can be directly applied to predict the power output of any of the considered configuration for the future wind farms.

4.1. Improved offshore wind generation modelling within sequential Monte-Carlo for adequacy studies

The offshore wind generation model is composed of three main parts, i.e., the wind model, the wind turbine availability model and the topology-aware ML surrogate.

Journal Pre-proof

For the wind model, when taking the wake effects into account, the wind direction has a significant impact on the output power of wind turbines. Moreover, when generating wind data, it is important to maintain the correlation between wind speeds and directions among the different neighbouring wind farms (prevailing winds may occur). To that end, we use a Vector Auto-Regressive Moving Average (VARMA) model, which augments the ability of ARMA models (that accurately represent time dependencies) with a representation of cross-variables correlations [51]. The interested reader can find detailed information regarding the VARMA model in Appendix A.

To account for wind turbine outages, the operational cycle of a wind turbine is represented through a two-state model using a probabilistic model similar to the one described in [10]. Turbine failures are fully encompassed by the developed ML surrogate, which can capture the redistribution of wind due to disabled turbines and the subsequent change in electrical power output.

Wind scenarios are then fed to the ML-based wind farm surrogate developed in section 3, along with turbine availabilities. Offshore wind power profiles are obtained, fully accounting for the time correlation between consecutive states and power losses due to wake interactions between turbines.

4.2. Offshore wind generation for the Belgian power system in 2030

Currently, 2.2 GW of offshore wind generation capacity is installed in the Belgian North Sea, in a zone consisting of 9 clustered wind farms [29]. However, aligned with the ambitious objectives of the energy transition, the Belgian government has decided to build a second offshore wind zone by 2030, but the layout of these future wind farms is still unknown. Therefore, it is important to consider multiple scenarios for this future offshore zone in order to account for the current uncertainty on the energy that will be provided by the different wind farms, as proposed in preliminary studies carried out by the Belgian government [52]. Indeed, there is a high uncertainty on how wind farm developers will optimize their installation (turbine location, turbine choice, power density). Therefore, 5 installed capacities in the second offshore zone are investigated: 1750, 2100, 2800, 3150, and 3500 MW. Moreover, different wind turbine technologies are also considered. As the trend in recent offshore wind installations is to invest in larger turbines, the following 3 turbines types are assessed: a 13 MW turbine with a rotor diameter D of

Journal Pre-proof

220 m, a 15 MW turbine with D = 236m, and a boosted 17 MW variant of the latter. The location of turbines within the future offshore farms is determined using a traditional micro-siting method, complying with inter-turbine distance, environmental and shipping-constraints [53]. Examples of such layouts are presented in Fig. 8.

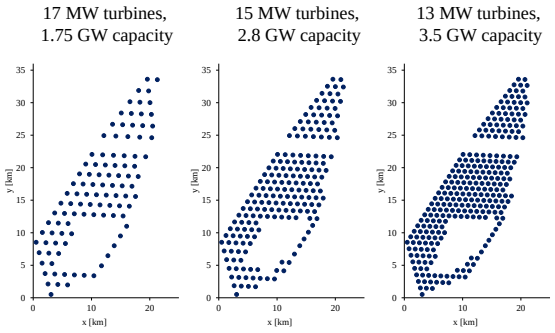


Figure 8: Second offshore zone (Princess Elizabeth) in the Belgian North Sea: examples of possible layouts

4.3. Numerical results

4.3.1. Sensitivity to the power density within wind farms

LOLE and LOEE values, as well as yearly offshore production, are reported in Table 5 for each combination of installed capacity and turbine model in the second offshore zone. As expected, when the installed power of offshore farms increases from 1.75 GW to 3.5 GW, the yearly produced energy grows as well. This, in turn, improves reliability indices, leading to a more robust supply of electricity. However, for a given turbine model, this effect becomes less significant as power density increases. The diminishing marginal returns can be explained by the fact that adding more turbines amplifies wake losses. Consequently, the installed capacity should thus result from a well-studied compromise between increasing the yearly offshore production while limiting power losses from wake effects.

4.3.2. Sensitivity to the choice of wind turbines

From Table 5, another interesting observation is that the choice of turbine model has a significant impact on adequacy indices. For example, for a 1.75 GW capacity in the second offshore zone, LOLE values are respectively

Journal Pre-proof

	LOLE	LOEE	Yearly offshore production
Case 1: 1.75 GW in the 2^{nd} offshore zone (6.2 MW/km²)			
13 MW turbines	29.54 h/y	36.43 GWh/y	15.72 TWh
15 MW turbines	28.73 h/y	35.55 GWh/y	15.92 TWh
17 MW turbines	28.22 h/y	34.30 GWh/y	15.99 TWh
Case 2: 2.1 GW in the 2^{nd} offshore zone (7.5 MW/km²)			
13 MW turbines	27.42 h/y	33.32 GWh/y	17.05 TWh
15 MW turbines	27.14 h/y	33.19 GWh/y	17.21 TWh
17 MW turbines	26.70 h/y	32.22 GWh/y	17.40 TWh
Case 3: 2.8 GW in the 2^{nd} offshore zone (10 MW/km²)			
13 MW turbines	25.72 h/y	31.12 GWh/y	19.26 TWh
15 MW turbines	25.04 h/y	30.51 GWh/y	19.68 TWh
17 MW turbines	24.85 h/y	30.54 GWh/y	19.87 TWh
Case 4: 3.15 GW in the 2^{nd} offshore zone (11.2 MW/km²)			
13 MW turbines	24.44 h/y	29.81 GWh/y	20.30 TWh
15 MW turbines	24.25 h/y	29.41 GWh/y	20.69 TWh
17 MW turbines	24.10 h/y	29.51 GWh/y	20.94 TWh
Case 5: 3.5 GW in the 2^{nd} offshore zone (12.5 MW/km²)			
13 MW turbines	24.42 h/y	30.10 GWh/y	21.33 TWh
15 MW turbines	24.17 h/y	29.59 GWh/y	21.80 TWh
17 MW turbines	23.48 h/y	28.40 GWh/y	22.10 TWh

Table 5: Adequacy indices, for different wind farm layouts in the second offshore zone.

29.54, 28.73 and 28.22 hours/year for 13, 15, and 17 MW turbines. LOEE values decrease from 36.43 GWh/year to 34.30 GWh/year when the rated power of turbines increases. However, the cost of larger turbines is higher, and wind farm operators might choose to install less powerful turbines if the difference in yearly energy does not justify the expense.

4.4. Computation time

As a reminder, since the ML proxy is independent of the wind farm configuration, the training phase is carried out only **once**, and the resulting ML proxy can be applied for any wind farm layout in the second offshore zone. Computations were performed on an Intel Core i7-9750 CPU (6 cores, 2.6 GHz). The whole training process takes approximately 465 minutes (less than 8 hours), including (1) carry out wind farm simulations to build the training database (12 minutes), (2) search for optimal hyperparam-

Journal Pre-proof

eters (442 minutes), and (3) train the models (11 minutes). In comparison, the entire training process for the farm-specific models (such as the ones developed in [10]), takes around 157 minutes for a single wind farm. In the proposed case study, training 24 models would have been necessary, resulting in over 62 hours of computational effort. Moreover, exploring an additional layout scenario for the second offshore zone would necessitate developing yet another model, thereby further extending the simulation duration. This clearly shows the advantage of the proposed topology-aware ML surrogate over the conventional approach that consists in building multiple farm-specific models.

Regarding the time needed to carry out Monte-Carlo simulations for each case, the average number of simulated years needed to reach convergence was around 500 to 700 Monte-Carlo years (each year requires yearly offshore power time series of 8760 samples). For the case 5 in Table 5 (3.5 GW in the second offshore zone with 15 MW turbines), one Monte-Carlo simulation would take between 412 and 577 hours with Floris-based simulations, while it only required 13 to 18 hours using the proposed generic ML model. The difference in computation time is huge, especially if one wants to consider several scenarios for the power system.

5. Conclusion

This paper introduces a ML-based surrogate model that efficiently captures aerodynamic losses in large offshore wind farms without compromising computational tractability within iterative power systems computations. The model is trained on a large database that is built using wind farm simulations. In the absence of real-life measurements data, wake models are leveraged to simulate wind farm power production in various wind situations. The ML surrogate is fed by new geometric and physics-informed input features such that our developed model is generic and topology-aware, i.e., it can be applied to any offshore wind farm. Moreover, the surrogate allows to account for disabled turbines (failures or maintenance) and the subsequent wake redistribution. The proposed methodology is easy to implement and to reproduce, as it only uses publicly available data and simulation outputs.

Four standard supervised ML algorithms are compared for the training process, and their performance is compared in terms of computational time, accuracy and model complexity. All ML models show strong performance

Journal Pre-proof

regarding both errors and inference time, but neural networks (more specifically multi-layer perceptrons) are chosen as they show the best compromise between accuracy and speed. The computation time is detailed, and compared against wind farm simulations: the general surrogate is shown to be more than 25 times faster than the simulations carried out with Floris. Our surrogate is also benchmarked against other well-established models based on machine learning and exhibiting capabilities to adapt to unseen wind farm configurations. Moreover, scalability and generalization to irregular layouts is also studied. It was shown that even for large wind farms with an irregular layout (unseen in the training database), error metrics and inference time remain low.

The model can then be directly integrated in problems involving iterative computations, e.g., adequacy studies using sequential Monte-Carlo simulations. Results of a test case, based on a simplified model of Belgium's 2030 power system, demonstrate the generalization abilities allow to account for various possibilities related to the uncertain topology of future wind farms. Indeed, with the topology-aware ML surrogate, the same model can be used to consider many possible farm configurations, thereby preventing the need to train one model per configuration, as it would be the case with traditional farm-specific surrogates. Outcomes show that increasing power density improves reliability indices, but diminishing marginal returns are observed after a certain point, due to the amplification of wake losses caused by the additional turbines. Moreover, the choice of turbine technology also impacts adequacy results, as larger rotor diameters decreases adequacy indices, but at the expense of higher turbine costs.

As a perspective, the generic ML model offers promising potential for optimizing the topology of future wind farms. Indeed, the developed topology-aware ML surrogate is especially appropriate for a wind farm layout optimization problem, where turbines are moved within the farm boundaries at each optimization step. Therefore, a different layout is seen at each iteration, justifying the need for a topology-aware model, applicable to any wind farm configuration.

Acknowledgments

ERA5 reanalysis is available from C3S Climate Data Store, see `https://doi.org/10.24381/cds.adbb2d47` (C3S, 2017).

Journal Pre-proof

This research is supported by the energy transition funds project "PhairywinD" organized by the Belgian FPS economy.

Appendix A. VARMA model for the generation of correlated time series of wind speed and wind direction.

In Auto-Regressive Moving Average (ARMA) models for wind speed, each value in the simulated time series depends on its own lagged values (AR part) but also on current and various past values of a stochastic term (MA part). The model is usually referred to as an ARMA(p, q) model where p and q are respectively the order of the auto-regressive and the moving average parts. First, the series y_t (the wind speed) is normalized to ensure stationarity, as the hourly wind speed distribution is non stationary due to the daily cycle and seasonality:

$$y_t = \frac{OW_t - \mu_t}{\sigma_t} \tag{A.1}$$

where t is an hour in the year, OW_t is the observed wind speed at hour t, μ_t is the mean of all observed wind speed at hour t over several years, and σ_t is the standard deviation of all observed wind speed at hour t.

Then, the data series y_t can be used to build the following ARMA(p, q) wind speed time series model [54]:

$$\begin{aligned} y_t = {} & \phi_1 * y_{t-1} + \phi_2 * y_{t-2} + ... + \phi_p * y_{t-p} \\ & + \alpha_t - \theta_1 * \alpha_{t-1} - \theta_2 * \alpha_{t-2} - ... - \theta_q * \alpha_{t-q} \end{aligned} \tag{A.2}$$

where ϕ_i (i = 1, 2, ..., p) and θ_j (j = 1, 2, ..., q) are respectively the auto-regressive and moving average parameters of the model, α_t is a normal white noise process with zero mean and variance σ_a^2, i.e., $\alpha_t \in NID(0, \sigma_a^2)$ with NID denoting Normally Independently Distributed. The maximum likelihood approach is adopted to estimate the values of ϕ_i, θ_j and σ_a^2. A grid search procedure based on the F-criterion is used to determine the order of the ARMA(p, q) model.

Once the wind speed time series model is established, the simulated wind speed can be calculated as:

$$SW_t = \mu_\tau + \sigma_\tau * y_t \tag{A.3}$$

Journal Pre-proof

VARMA(p, q) models for simulating correlated wind speeds and wind directions are just a generalization of ARMA(p, q) models, where the wind speed not only depends on its own lagged values, but also on the lagged values of wind directions. The same goes for the wind direction time series, and the wind vector is written as:

$$\begin{aligned} y_t = \Phi_1 * y_{t-1} + \Phi_2 * y_{t-2} + ... + \Phi_p * y_{t-p} \\ + \alpha_t - \Theta_1 * \alpha_{t-1} - ... - \Theta_q * \alpha_{t-q} \end{aligned} \tag{A.4}$$

where $y_t = [y_s, y_d]^t$ contains the data series corresponding to wind speed and wind direction and Φ_i and Θ_j are matrices of dimensions [2x2]. The methodology for estimating the parameters and for choosing the (p, q) order is the same as for ARMA models (F-criterion).

Four years (2018 to 2021) of wind speed and wind direction data from the ERA5 database are used to build the VARMA model. The F-criterion leads to a VARMA(3, 2) model. As a precautionary measure, a simulation procedure is needed to check whether the wind model can preserve the main characteristics of wind speed: the statistical properties of the simulated wind speed (mean, standard deviation, monthly distribution) should be compared with those obtained from the actual wind speed. In Fig. A.9, it can be seen that the VARMA model keeps the seasonality and monthly distribution of the wind speed. Moreover, the Pearson correlation coefficient is computed between the wind speed and the wind direction. For the observed data (2018-2021 from ERA5), the coefficient is 0.154, while the average coefficient is 0.135 for simulated data.

Journal Pre-proof

Journal Pre-proof

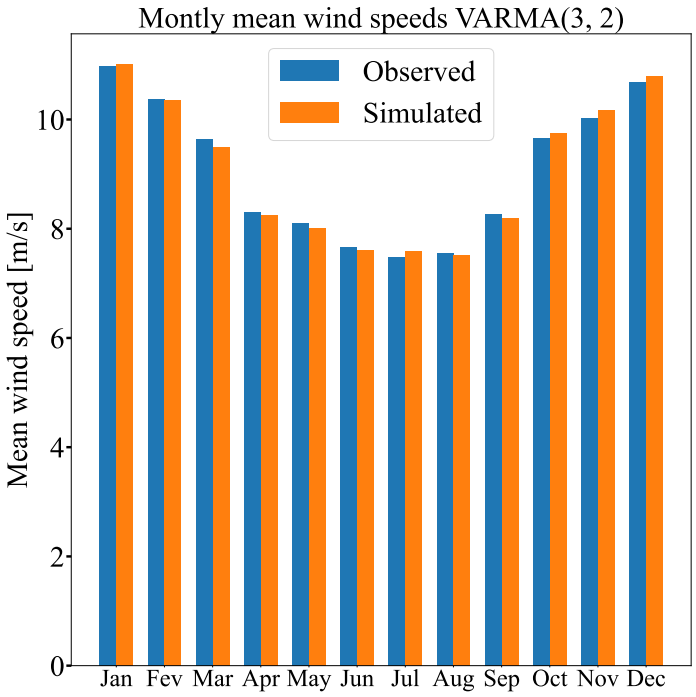


Figure A.9: Comparison of monthly wind speed between observed and simulated data with the VARMA(3,2) model.

References

[1] Global wind report, Tech. rep., Global Wind Energy Council (2024).

[2] D. Villanueva, A. Feijóo, J. L. Pazos, Simulation of correlated wind speed data for economic dispatch evaluation, IEEE Transactions on Sustainable Energy 3 (1) (2011) 142–149.

[3] Elia, Adequacy and flexibility study for Belgium 2024-2034, Tech. rep., Elia (2023).

[4] L. Söder, E. Tómasson, A. Estanqueiro, D. Flynn, B.-M. Hodge, J. Kiviluoma, M. Korpås, E. Neau, A. Couto, D. Pudjianto, et al., Review of wind generation within adequacy calculations and capacity markets for different power systems, Renewable and Sustainable Energy Reviews 119 (2020) 109540.

Journal Pre-proof

[5] A. Almutairi, M. H. Ahmed, M. Salama, Use of MCMC to incorporate a wind power model for the evaluation of generating capacity adequacy, Electric Power Systems Research 133 (2016) 63–70.

[6] H. Kim, C. Singh, A. Sprintson, Simulation and estimation of reliability in a wind farm considering the wake effect, IEEE Transactions on Sustainable Energy 3 (2) (2012) 274–282.

[7] F. Pelletier, C. Masson, A. Tahan, Wind turbine power curve modelling using artificial neural network, Renewable Energy 89 (2016) 207–214.

[8] M. Schlechtingen, I. F. Santos, S. Achiche, Using data-mining approaches for wind turbine power curve monitoring: a comparative study, IEEE Transactions on Sustainable Energy 4 (3) (2013) 671–679.

[9] H. Zhou, Y. Qiu, Y. Feng, J. Liu, Power prediction of wind turbine in the wake using hybrid physical process and machine learning models, Renewable Energy 198 (2022) 568–586.

[10] T.-H. Nguyen, J.-F. Toubeau, E. D. Jaeger, F. Vallee, Fast and reliable modelling of offshore wind generation for adequacy studies, IEEE Transactions on Industry Applications (2023) 1–10doi:10.1109/TIA.2023.3307076.

[11] Z. Ti, X. W. Deng, M. Zhang, Artificial neural networks based wake model for power prediction of wind farm, Renewable Energy 172 (2021) 618–631.

[12] N. Zehtabiyan-Rezaie, A. Iosifidis, M. Abkar, Physics-guided machine learning for wind-farm power prediction: Toward interpretability and generalizability, PRX Energy 2 (1) (2023) 013009.

[13] J. Park, J. Park, Physics-induced graph neural network: An application to wind-farm power estimation, Energy 187 (2019).

[14] A. Kusiak, Renewables: Share data on wind energy, Nature 529 (7584) (2016) 19–21.

[15] M. Kretschmer, J. Jonkman, V. Pettas, P. W. Cheng, Fast. farm load validation for single wake situations at alpha ventus, Wind Energy Science 6 (5) (2021) 1247–1262.

Journal Pre-proof

[16] National Renewable Energy Lab (NREL), FLORIS v.4.1 (2024).
URL https://github.com/NREL/floris

[17] M. M. Pedersen, A. Meyer Forsting, P. van der Laan, R. Riva, L. A. Alcayaga Romàn, J. Criado Risco, M. Friis-Møller, J. Quick, J. P. Schøler Christiansen, R. Valotta Rodrigues, B. T. Olsen, P.-E. Réthoré, Pywake 2.5.0: An open-source wind farm simulation tool (2023).
URL https://gitlab.windenergy.dtu.dk/TOPFARM/PyWake

[18] J. Schmidt, L. Vollmer, M. Dörenkämper, B. Stoevesandt, Foxes: Farm optimization and extended yield evaluation software, Journal of Open Source Software 8 (86) (2023) 5464. doi:10.21105/joss.05464.
URL https://doi.org/10.21105/joss.05464

[19] R. Mudafort, J. Quick, J. Schulte, Comparison of steady-state analytical wake models implemented in wind farm analysis software, in: Journal of Physics: Conference Series, Vol. 2767, IOP Publishing, 2024.

[20] M. T. van Beek, A. Viré, S. J. Andersen, Sensitivity and uncertainty of the Floris model applied on the Lillgrund wind farm, Energies 14 (5) (2021) 1293.

[21] B. M. Doekemeijer, D. van der Hoek, J.-W. van Wingerden, Closed-loop model-based wind farm control using FLORIS under time-varying inflow conditions, Renewable Energy 156 (2020) 719–730.

[22] A. Niayifar, F. Porté-Agel, Analytical modeling of wind farms: A new approach for power prediction, Energies 9 (9) (2016) 741.

[23] L. P. Chamorro, F. Porté-Agel, A wind-tunnel investigation of wind-turbine wakes: boundary-layer turbulence effects, Boundary-layer meteorology 132 (2009) 129–149.

[24] W. Zhang, C. D. Markfort, F. Porté-Agel, Wind-turbine wakes in a convective boundary layer: A wind-tunnel study, Boundary-layer meteorology 146 (2013) 161–179.

[25] Y.-T. Wu, F. Porté-Agel, Atmospheric turbulence effects on wind-turbine wakes: An LES study, energies 5 (12) (2012) 5340–5362.

Journal Pre-proof

[26] M. Gaumond, P.-E. Réthoré, S. Ott, A. Pena, A. Bechmann, K. S. Hansen, Evaluation of the wind direction uncertainty and its impact on wake modeling at the horns rev offshore wind farm, Wind Energy 17 (8) (2014) 1169–1178.

[27] A. Crespo, J. Herna, et al., Turbulence characteristics in wind-turbine wakes, Journal of wind engineering and industrial aerodynamics 61 (1) (1996) 71–85.

[28] J. Schneemann, F. Theuer, A. Rott, M. Dörenkämper, M. Kühn, Offshore wind farm global blockage measured with scanning lidar, Wind Energy Science 6 (2) (2021) 521–538.

[29] Offshore wind farms in belgium, Last access: June 2024.
URL https://odnature.naturalsciences.be/mumm/en/windfarms/

[30] RAVE: Research at Ahlpa Ventus, last access: June 2024.
URL https://rave-offshore.de/en/data.html

[31] H. Hersbach, et al., ERA5 hourly data on single levels from 1959 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS), Last access: June, 2024 (2018).
URL 10.24381/cds.adbb2d47

[32] Technical Committee 88 Wind energy generation systems, IEC 61400-3-1, Tech. rep., International Electrotechnical Commission (IEC) (2019).

[33] R. J. Barthelmie, et al., Modelling and measuring flow and wind turbine wakes in large wind farms offshore, Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology 12 (5) (2009) 431–444.

[34] N. S. Ghaisas, C. L. Archer, Geometry-based models for studying the effects of wind farm layout, Journal of Atmospheric and Oceanic Technology 33 (3) (2016) 481–501.

[35] S. Xie, C. Archer, Self-similarity and turbulence characteristics of wind turbine wakes via large-eddy simulation, Wind Energy 18 (10) (2015) 1815–1838.

Journal Pre-proof

[36] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, L. Yang, Physics-informed machine learning, Nature Reviews Physics 3 (6) (2021) 422–440.

[37] B. Huang, J. Wang, Applications of physics-informed neural networks in power systems: a review, IEEE Transactions on Power Systems 38 (1) (2022) 572–588.

[38] D. Vahidi, F. Porté-Agel, A physics-based model for wind turbine wake expansion in the atmospheric boundary layer, Journal of Fluid Mechanics 943 (2022) A49.

[39] N. O. Jensen, A note on wind generator interaction, Vol. 2411, 1983.

[40] R. J. Barthelmie, S. T. Frandsen, M. Nielsen, S. Pryor, P.-E. Rethore, H. E. Jørgensen, Modelling and measurements of power losses and turbulence intensity in wind turbine wakes at Middelgrunden offshore wind farm, Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology 10 (6) (2007) 517–528.

[41] D. Bertoin, J. Bolte, S. Gerchinovitz, E. Pauwels, Numerical influence of ReLU'(0) on backpropagation, Advances in Neural Information Processing Systems 34 (2021) 468–479.

[42] T.-H. Nguyen, Improved modelling of offshore wind generation using machine learning - on a reliable supply of electricity in future power systems, Ph.D. thesis, University of Mons (2025).

[43] C. Yan, Y. Pan, C. L. Archer, A general method to estimate wind farm power using artificial neural networks, Wind Energy 22 (11) (2019) 1421–1432.

[44] G. Duthé, F. de Nolasco Santos, I. Abdallah, P.-É. Réthore, W. Weijtjens, E. Chatzi, C. Devriendt, Local flow and loads estimation on wake-affected wind turbines using graph neural networks and pywake, in: Journal of Physics: Conference Series, Vol. 2505, IOP Publishing, 2023.

[45] R. Valotta Rodrigues, M. M. Pedersen, J. P. Schøler, J. Quick, P.-E. Réthoré, Speeding up large-wind-farm layout optimization using gradients, parallelization, and a heuristic algorithm for the initial layout, Wind Energy Science 9 (2) (2024) 321–341.

Journal Pre-proof

[46] ENTSO-E, ENTSO-E Transparency Platform, last access: April 2025. URL https://transparency.entsoe.eu/

[47] R. Billinton, W. Li, Reliability Assessment of Electric Power Systems Using Monte Carlo Methods, Springer, 2013.

[48] A. A. Kadhem, N. I. A. Wahab, I. Aris, J. Jasni, A. N. Abdalla, Computational techniques for assessing the reliability and sustainability of electrical power systems: A review, Renewable and Sustainable Energy Reviews 80 (2017) 1175–1186.

[49] F. Vallée, J. Lobry, O. Deblecker, System reliability assessment method for wind power integration, IEEE Transactions on Power Systems 23 (3) (2008) 1288–1297.

[50] R. N. Allan, R. Billinton, N. Abdel-Gawad, The IEEE reliability test system-extensions to and evaluation of the generating system, IEEE Transactions on Power Systems 1 (4) (1986) 1–7.

[51] E. Erdem, J. Shi, ARMA based approaches for forecasting the tuple of wind speed and direction, Applied Energy 88 (4) (2011) 1405–1414.

[52] 3E, LCOE offshore wind in the Princess Elizabeth zone: Analysis of various density scenarios, Tech. rep., FOD Economie (2021).

[53] W. Munters, B. Adiloglu, S. Buckingham, J. van Beeck, Wake impact of constructing a new offshore wind farm zone on an existing downwind cluster: a case study of the belgian princess elisabeth zone using floris, in: Journal of Physics: Conference Series, Vol. 2265, IOP Publishing, 2022.

[54] R. Billinton, H. Chen, R. Ghajar, Time-series models for reliability evaluation of power systems including wind energy, Microelectronics Reliability 36 (9) (1996) 1253–1261.

Journal Pre-proof

Journal Pre-proof

Topology-Aware Surrogate for Future Offshore Wind Farms Using Machine Learning

Thuy-Hai Nguyen[a,*], Jean-François Toubeau[a], Emmanuel De Jaeger[b], François Vallée[a]

[a] *Electrical Power Engineering Unit, University of Mons, 31 Boulevard Dolez, Mons, 7000, Belgium*
[b] *Department of Mechatronics, Electrical Energy and Dynamic Systems, University of Louvain, 2 Place du Levant, Louvain-la-Neuve, 1348, Belgium*

Abstract

As the role of offshore wind generation increases in modern power systems, the need for enhanced modelling techniques becomes critical. In problems involving iterative computations, simplified offshore wind farm models often disregard complex aerodynamic effects in order to reduce computational time. However, those effects have a significant impact on the produced electricity. To address this current modelling limitation, we develop a new topology-aware wind farm surrogate using supervised Machine Learning techniques. This surrogate model is trained to capture the complex relationship between free flow wind information and electrical output, irrespective of the wind farm layout. This allows bypassing the need to train a different model for each wind farm, while enabling the simulation of farms lacking historical data (such as those in construction phase). This is achieved by enriching the input space of the Machine Learning surrogate with novel geometric and physics-informed features, making it adaptable to any offshore wind farm configuration.

Keywords: Machine Learning, Offshore wind, Surrogate, Topology-aware, Wake effects

*Corresponding author
Email address: thuy-hai.nguyen@umons.ac.be

Journal Pre-proof

Journal Pre-proof

- Machine learning is used to develop a surrogate simulating offshore wind farms
- Our model is topology-aware, i.e., able to generalize to unseen wind farm layouts
- The surrogate is applicable to any offshore wind farm, even farms yet to be built
- Wake losses are accurately modelled while maintaining a low computational time
- The surrogate, trained once, can account for the uncertainty of future wind farms

Journal Pre-proof

Journal Pre-proof

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof