PAPER • **OPEN ACCESS**

Surrogate-assisted optimization of floating wind turbine substructure

Journal of Physics: Conference Series **2767** (2024) 062032 doi:10.1088/1742-6596/2767/6/062032

Surrogate-assisted optimization of floating wind turbine substructure

M Baudino Bessone[1], D Singh[1], T Kalimeris[2], E Bachynski-Polić[3], A Viré[1]

[1]Aerospace Engineering, TU Delft, Kluyverweg 1, 2629 HS Delft, The Netherlands
[2]Temporary Works Design, Marconistraat 16 3029 AK, Rotterdam, The Netherlands
[3]Department of Marine Technology, NTNU, Jonsvannsveien 82, 7050, Trondheim, Norway

E-mail: m.baudinobessone@tudelft.nl

Abstract. This paper presents a surrogate-assisted optimisation approach to speed up the substructure analysis in the preliminary design phase. The approach consists of replacing the radiation-diffraction analysis in a frequency domain analysis model for floating wind turbines with a data-driven surrogate model predicting the hydrodynamic coefficients for parameterised substructure geometries. This procedure is compared with the reference approach of estimating the hydrodynamic coefficients via radiation-diffraction analysis. A representative use case of assessing the trade-off between minimising the capital cost and reducing the wave-induced nacelle acceleration standard deviation for a semi-submersible substructure is presented. The accuracy of the surrogate model is found to increase significantly up to training datasets consisting of 400 designs and less noticeably afterwards. For a dataset consisting of 400 designs, the mean error on the prediction of the hydrodynamic coefficients and the error at one standard deviation from the mean are generally below 7% and 10%, respectively. For the same dataset size, the mean error on the most probable maximum wave-induced pitch over a 3h storm period is below 17%, while the error at one standard deviation from the mean is lower than 27%. The same values for the most probable maximum nacelle acceleration are under 7% and 12%, respectively. The surrogate model can capture the trade-off between the two objective functions, and the optimal designs identified with the surrogate model generally follow the same trend as those obtained with the reference model. However, relying on the surrogate model for performing the analysis of the substructure introduces local minima in the objective function that cause a discrepancy between the optimal designs identified with the surrogate model and those identified with the reference model.

1. Introduction

Floating wind is a promising renewable energy source, enabling the deployment of wind farms in regions where conventional bottom-fixed substructures are no longer economically appealing. However, a significant reduction in the levelized cost of energy (LCoE) of floating wind systems is necessary to make it competitive with other energy sources. The substructure capital cost constitutes a significant part of the life cycle cost for a floating wind farm. This consideration has driven a wide interest in adopting various optimisation algorithms to minimise the production cost for different floating substructure concepts [1–3]. The floating unit analysis is carried out in the frequency or time domain. Frequency domain approaches are usually adopted in the preliminary design stage, as they allow quicker design space exploration than time domain models. However, even adopting frequency domain analysis, the substructure optimisation

becomes time-consuming when the floater first-order hydrodynamic coefficients are obtained via radiation-diffraction analysis software. Therefore, researchers developed methods to avoid running radiation-diffraction analysis within the substructure optimisation workflow, such as interpolating pre-computed, gridded values [4], or decomposing complex substructures into the constituting elements, followed by superposition of the hydrodynamic coefficients obtained for these simpler geometries [5].

In this work, we present a novel surrogate-assisted optimisation approach for the design of floating wind substructures. The ground truth is based on the open-source frequency domain analysis model for floating wind units, RAFT [6], developed by NREL, where the radiation-diffraction analysis is performed via the open-source software HAMS [7] for a given substructure geometry. In our approach, the design optimisation problem is accelerated through the use of state-of-the-art data-driven models, where the surrogate model replaces HAMS in RAFT's workflow. The hydrodynamic coefficients estimated by the surrogate model are then used within RAFT to perform the substructure analysis in the frequency domain. This procedure is compared with the ground truth for the representative use case of assessing the trade-off between minimising the substructure capital cost and reducing the waves-induced nacelle acceleration.

2. Floater design parameters

The floating substructure selected for this study is a semi-submersible whose geometry is reported in Figure 1. Three design variables are defined, which are represented in Figure 1 and are reported in Table 1 together with the bounds. The semi-submersible's remaining geometrical and structural design parameters are defined as constants or fully determined by combinations of other parameters and design variables and reported in Table 2. The ballast mass is composed of 70% solid concrete, with density 2400kg/m^3, and of 30% water, with density 1025kg/m^3.

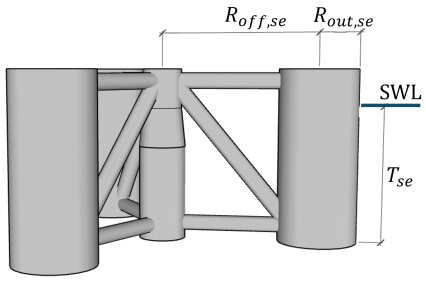

Figure 1: Semi-submersible geometry and design variables. The figure is not drawn to scale.

Table 1: Semi-submersible design variables.

Design variable	Symbol	Bounds
Radius of the outer columns	$R_{out,se}$	2.7m-6.2m
Radial distance between the central and outer columns	$R_{off,se}$	40m-70m
Draft between the still water level, SWL, and the keel	T_{se}	15m-45m

3. Machine learning framework

3.1. Surrogate model

Data-driven surrogate modelling is a widely used approach for estimating the underlying process of a complex system using information derived solely from the observed input-output data pairs. A database with n samples is defined as $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}$, where $\boldsymbol{x} \in \mathbb{R}^m$ are m-dimensional input features of the data sample i and $y_i \in \mathbb{R}$ is the corresponding response. The *training* process consists of estimating an approximation of the underlying function that maps $\boldsymbol{x} \rightarrow y$ by minimising the expected value of a user-defined objective function $\mathcal{L}$ over the joint distribution of the training subspace.

In this study, surrogate modelling is performed using XGBoost [10], a widely used and very powerful machine learning approach that falls under the category of tree-based ensemble

Table 2: Summary of semi-submersible geometrical and structural parameters.

Parameter	Value	Parameter	Value
Freeboard height	10m	Thickness of the central and outer columns	0.05m [1]
Depth of the taper of the central column	Twice the 50 years significant wave amplitude, $\zeta_{s,50} = 5.4$m	Central column taper angle	10.3° [8]
Connection height of the cross braces with the upper section of the central column	75% of freeboard height [9]	Connection depth of the cross braces with the lower section of the outer column	80% of draft [9]
Diameter of cross braces, upper and lower pontoons	As proposed by Karimi et al., [1]	Cross braces and pontoons wall thickness to radius ratio	0.022 [9]

methods. Ensembling works by estimating the best-fitting underlying functions f on randomly picked subsets of the training dataset and input features and building the final predictive model as a combination or ensemble of the learned functions. XGBoost uses an additive ensemble, wherein the predicted value at any i^{th} sample, $\hat{y}_i$ is calculated from K additive functions - each function f_k being calculated by an individual regression tree in the space of regression trees, $\mathcal{F}$.

$$\hat{y}_i = \sum_{k=1}^{K} f_k(\boldsymbol{x}_i), f_k \in \mathcal{F}, \tag{1}$$

A canonical representation of a binary regression tree can be seen in Figure 2. Each internal node splits either into two leaves or another internal node and a leaf. The decision in an internal node is based on a split s of one of the feature values. The decision rules defining the tree structure are represented by q. The number of leaves is represented by T, and the weight or score assigned to each leaf is denoted by w. Therefore, the space of trees can be defined by $\mathcal{F} = f(\boldsymbol{x}) = w_{q(x)}(q : \mathbb{R}^m \rightarrow T, w \in \mathbb{R}^T)$. For an input condition $\boldsymbol{x}_i$, a series of decisions are made for each tree k, until a leaf is encountered. The value of $f_k(\boldsymbol{x}_i)$ is then equal to the score of the leaf the decision sequence ends on. The final value of the prediction $\hat{y}_i$ is calculated by summing the scores using Equation (1) over all K trees.
The regularized objective function used to determine optimal tree functions is defined as:

$$Obj = \sum_i l(\hat{y}_i, y_i) + \sum \Omega(f_k), \tag{2}$$

where $\Omega(f) = \gamma T + 0.5\lambda||w||$ is a regularisation function that prevents overfitting by penalising the number of leaves through the hyperparameter γ and by encouraging the weights to be small through the hyperparameter λ. l is a user-defined differentiable, convex loss function. In this study, l is the squared error between y_i and $\hat{y}_i$.

The gradients of Equation (2) are calculated using Taylor series expansion. However, the objective is a function of trees, and finding trees that minimise the objective along a suitable gradient is one of the main problems of tree based ensemble approaches. XGBoost addresses the issue using *gradient boosting*, which is an algorithm that combines weak learners (in the form of individual decision trees) to form a single strong learner in a greedy and iterative fashion. In summary, learning the optimal function f_k can be divided essentially into two problems:

- Finding the leaf values w: For a fixed tree structure, w can be analytically calculated based on the first and second gradients of the objective function [10].
- Finding the tree structure q: XGBoost employs a greedy algorithm that iteratively adds leaves to the tree. The tree is initialised by a single leaf, which is transformed into an internal node with two leaves on the next iteration. For each feature, the split is linearly scanned, and the best value is determined based on the value of the objective function. The new tree structure is preserved if there is a net information gain. The process is repeated until the user-defined value of maximum tree depth is reached. This step is the most computationally expensive as it requires linear scanning of the values of all the features. Further, the tree is pruned to simplify the tree structure.

In this project, we implement a multi-output tree structure, which uses vector leaves to preserve the correlation between the output columns during modelling. This feature is extremely novel in XGBoost and still subject to extensive research and testing. The reason for using a multi-output tree structure will be made clear in the following section.

3.2. Hyperparameters

The hyperparameters chosen for this particular study are listed in Table 3.

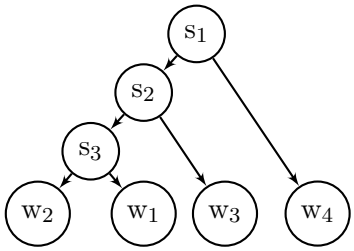

Figure 2: A canonical representation of a binary decision tree showing the internal nodes with split (s) and the leaves with weights (w).

Table 3: Hyperparameter values for the XGBoost model.

Hyperparameter	**Value**
Number of trees	150
Learning rate	0.1
Max. tree depth	10
Strategy	Multi-output tree
Loss function	Squared error

3.3. Training framework

The input features are the design variables listed in Section 2. Sobol sampling is used to perform quasi-random uniform sampling within the boundaries of the design space. The use case is simplified to only consider aligned wind and waves from a single direction, such that only the surge, heave and pitch degrees of freedom of the floater are considered. Furthermore, only the vertical columns of the semi-submersible are considered for radiation-diffraction analysis, while strip-theory is used for the slender pontoons and cross-braces. The frequency-dependent output parameters include:

- Hydrodynamic added mass: surge, A_{11}, heave, A_{33}, pitch, A_{55}, coupling term surge-pitch A_{51}
- Wave radiation damping: surge, B_{11}, heave, B_{33}, pitch, B_{55}, coupling term surge-pitch, B_{51}
- Wave excitation forces in surge, heave, and pitch

For each output parameter, we train a separate XGBoost model. The output columns, for instance, for the A_{11} element, are the values of A_{11} from f_{min} =0.004 Hz to f_{max} =0.4 Hz with a step of 0.004 Hz, resulting in 100 output columns per element.

The multi-output tree structure in XGBoost takes into account the correlation of the matrix element values across the frequency range, thereby predicting smooth continuous functions of frequency.

4. Case study

The surrogate model is applied to the problem of assessing the trade-off between minimising the capital cost of a semi-submersible substructure sized for the 5MW NREL reference wind turbine and reducing the standard deviation of the nacelle acceleration from pitch and surge, which is regarded as an indicator of floating wind turbine performance [1]. The optimisation problem is formally stated as:

Minimise

$$w_{C_{se}} C_{se} + w_{\sigma_{a,w}} \sigma_{a,w} \tag{3}$$

Subject to

$$T_{se} \geq T_{se,min} \tag{4}$$

$$\theta_{st} \leq \theta_{st,max} \tag{5}$$

$$\theta_{tot,n} \leq \theta_{tot,max} \tag{6}$$

$$a_{dyn,n} \leq a_{dyn,max} \tag{7}$$

In the objective function, C_{se} is the adimensional capital cost of the substructure, expressed as:

$$C_{se} = [M_s C_s (1 + MCF) + M_c C_c]/C_{max} \tag{8}$$

M_s and M_c are the mass of the steel substructure and concrete in the ballast. C_s is the steel price, 1000€/t, and C_c is the concrete cost, 86.6€/t. MCF is the Manufacturing Complexity Factor accounting for the semi-submersible fabrication cost, which is computed as the average of the values provided in the literature for steel semi-submersible with braces, 2.35 [11–13]. C_{max} is the cost of the semi-submersible when all the design variables are maximised.

$\sigma_{a,w}$ is the adimensional weighted sum of the standard deviation of nacelle acceleration across all the environmental conditions considered for the analysis, which are listed in Table 4. $\sigma_{a,w}$ is expressed as:

$$\sigma_{a,w} = \sum_{n=1}^{N_{env}} p_n \sigma_{a,n}/a_{dyn,max} \tag{9}$$

p_n is the probability of occurrence of the n^{th} environmental conditions. $\sigma_{a,n}$ is the standard deviation of nacelle acceleration resulting from wave loads computed for the n^{th} environmental condition. $a_{dyn,max}$ is the maximum allowable nacelle acceleration, $0.2g$, with g gravitational acceleration of 9.81 m/s^2. $w_{C_{se}}$ and $w_{\sigma_{a,w}}$ are two weights adopted to vary the relative

Table 4: Environmental conditions (EC) for substructure analysis. Wind speed Ws, significant wave height, H_s, peak period T_p, and probability of occurrence p for Buchan Deep site [15].

EC	Unit	1	2	3	4	5	6	7	8	9	10	11	12
Ws	m/s	3.0	5.0	7.0	9.0	11.0	13.0	15.0	17.0	19.0	21.0	23.0	25.0
H_s	m	0.92	1.09	1.29	1.52	1.80	2.12	2.50	2.96	3.49	4.13	4.87	5.76
T_p	s	6.7	6.9	7.1	7.40	7.7	8.0	8.4	8.80	9.2	9.7	10.3	11.0
p		7.6%	11.0%	13.8%	15.1%	14.7%	12.7%	9.4%	5.8%	3.7%	1.9%	0.9%	0.3%

importance of substructure capital cost and standard deviation of nacelle acceleration within the objective function, with $w_{C_{se}} + w_{\sigma_{a,w}} = 1$ and $w_{C_{se}}, w_{\sigma_{a,w}} \in [0, 1]$.

Equations 4 to 7 illustrate the constraints adopted in the optimisation. In Equation (4), $T_{se,min}$ is the minimum allowed draft for the substructure, expressed as in [14]. The maximum steady pitch at rated wind speed, θ_{st}, is enforced to be lower than the maximum allowed pitch, $\theta_{st,max}$, equal to 6^o, as expressed in Equation (5). The maximum pitch angle attained by the structure in each of the environmental conditions reported in Table 4, $\theta_{tot,n} = (\theta_{st,n} + \theta_{dyn,n})$, is enforced to be lower or equal than the maximum allowed pitch angle, $\theta_{tot,max}$, 8^o. $\theta_{st,n}$ is the steady pitch angle attained under the action of rotor thrust for environmental condition n, and $\theta_{dyn,n}$ is the maximum dynamic pitch angle attained by the structure due to wave loads, which is estimated as:

$$\theta_{dyn,n} = \sigma_{\theta,n} \sqrt{2 \ln \frac{T_{st}}{T_z}} \tag{10}$$

In this expression $\sigma_{\theta,n}$ is the standard deviation of the structure's pitch response due to wave loads, T_{st} is the storm length, 3h, and T_z is the average zero-crossing period. Finally, the wave-induced maximum nacelle acceleration from pitch and surge for each environmental condition, $a_{dyn,n}$, is constrained by its maximum allowable value, $a_{dyn,max}$. $a_{dyn,n}$ is computed similarly to $\theta_{dyn,n}$:

$$a_{dyn,n} = \sigma_{a,n} \sqrt{2 \ln \frac{T_{st}}{T_z}}, \tag{11}$$

The optimisation workflow is set in OpenMDAO [16], adopting the Constrained Optimisation by Linear Approximation (COBYLA) optimisation algorithm.

5. Results and discussion

5.1. Convergence

As the prediction accuracy of the data-driven models is driven by the number of training samples and the distribution of the training population, we perform a comprehensive convergence study. We test the performance of the surrogate model on different training sample designs ranging from 100 to 1800. For each subset, we perform 20-fold cross-validation, where we test the model on 100 test cases for each fold. The relative error described in Equation (12) is used as the accuracy metric for this study.

$$\varepsilon_{hyd} = \frac{\int_{f_{min}}^{f_{max}} |y_{ref} - \hat{y}| \, df}{\int_{f_{min}}^{f_{max}} |y_{ref}| \, df} \tag{12}$$

where y_{ref} and $\hat{y}$ are the hydrodynamic coefficients at each discrete frequency estimated through HAMS and the surrogate models, respectively.

For each training design size, the mean error, $\bar{\varepsilon}_{hyd}$ is defined as the average of the relative error ε_{hyd} over the 20 random data folds, averaged over the 100 test designs:

$$\bar{\varepsilon}_{hyd} = \mu_{n_{tests}}[\mu_{n_{seeds}}[\varepsilon_{hyd}]] \text{ and } \bar{\sigma}_{hyd} = \mu_{n_{tests}}[\sigma_{n_{seeds}}[\varepsilon_{hyd}]] \tag{13}$$

Similarly, the standard deviation of the error, $\bar{\sigma}_{hyd}$, is defined for each training dataset size as the standard deviation of ε_{hyd} over the twenty random datasets, averaged over the 100 test cases. $\bar{\varepsilon}_{hyd}$ and $\bar{\sigma}_{hyd}$ at the variation of training dataset size are reported in Figure 3 for the hydrodynamic added mass. In the Figure, the squares denote $\bar{\varepsilon}_{hyd}$, while the error bars denote $\bar{\sigma}_{hyd}$. $\bar{\varepsilon}_{hyd}$ and $\bar{\sigma}_{hyd}$ have also been computed for radiation damping and wave excitation load magnitude. However results have not been portrayed graphically for brevity, as they are well aligned to what can be observed for the hydrodynamic added mass. For all the hydrodynamic quantities under consideration, both the mean error and standard deviation of error decrease noticeably up to 400 training samples, whereas the decrease in both $\bar{\varepsilon}_{hyd}$ and

$\bar{\sigma}_{hyd}$ is less noticeable for training datasets with more than 400 samples. The mean error is below 7% for all the quantities predicted already for 400 samples training dataset, whereas it slightly exceeds 10% at one standard deviation for the mean only for A_{55}. Overall, the largest $\bar{\varepsilon}_{hyd}$ and $\bar{\sigma}_{hyd}$ are found for the pitch-related quantities, while the lowest mean error and error standard deviation are identified in the heave-related hydrodynamic coefficients. This is likely related to the fact that heave-related hydrodynamic coefficients for the chosen semi-submersible geometry vary mostly with $R_{out,se}$ and are less affected by $R_{off,se}$ and T_{se}. Instead, pitch-related quantities are significantly affected by all the semi-submersible design variables.

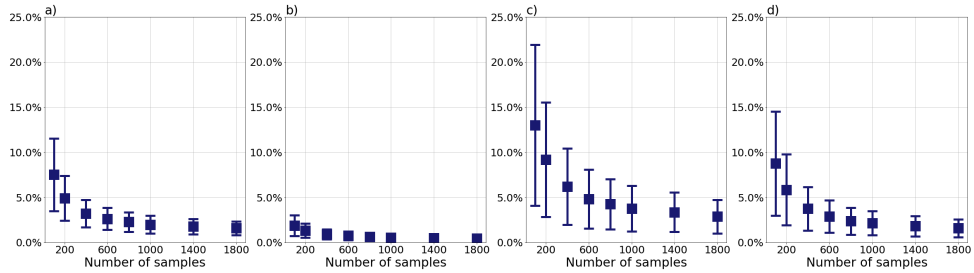

Figure 3: Convergence trend of a) A_{11}, b) A_{33}, c) A_{55}, and d) A_{51}.

For the application of the surrogate model to the optimisation use case, it is of interest to evaluate the influence of the error in the prediction of the hydrodynamic coefficients at the variation of the training dataset size on the optimisation constraints and objective function. Therefore, a second convergence study is carried out, comparing the maximum response in pitch $\hat{\theta}_{dyn,n}$ and maximum nacelle acceleration $\hat{a}_{dyn,n}$ due to wave loads predicted by RAFT by adopting the surrogate model, with the values that would be predicted by solving the diffraction-radiation problem with HAMS, $\theta_{dyn,n,ref}$ and $a_{dyn,n,ref}$, for each environmental condition in Table 4. The performance metrics for this comparison are the relative errors between these quantities:

$$\varepsilon_{\theta_{dyn,n}} = \frac{|\hat{\theta}_{dyn,n} - \theta_{dyn,n,ref}|}{|\theta_{dyn,n,ref}|} \text{ and } \varepsilon_{a_{dyn,n}} = \frac{|\hat{a}_{dyn,n} - a_{dyn,n,ref}|}{|a_{dyn,n,ref}|} \tag{14}$$

Similarly to the previous case, the mean error on the maximum wave-induced response in pitch, $\bar{\varepsilon}_{\theta_{dyn,n}}$, and the mean error on the maximum wave-induced nacelle acceleration, $\bar{\varepsilon}_{a_{dyn,n}}$, are defined for each training dataset size as the average of the relative error, respectively $\varepsilon_{\theta_{dyn,n}}$ and $\varepsilon_{a_{dyn,n}}$, over the twenty random data folds, averaged over the 100 test designs:

$$\bar{\varepsilon}_{\theta_{dyn,n}} = \mu_{n_{tests}}[\mu_{n_{seeds}}[\varepsilon_{\theta_{dyn,n}}]] \text{ and } \bar{\varepsilon}_{a_{dyn,n}} = \mu_{n_{tests}}[\mu_{n_{seeds}}[\varepsilon_{a_{dyn,n}}]] \tag{15}$$

The standard deviations of the error on the maximum wave-induced pitch $\bar{\sigma}_{\theta_{dyn,n}}$ and nacelle-acceleration $\bar{\sigma}_{a_{dyn,n}}$ are computed as the standard deviation of, respectively $\varepsilon_{\theta_{dyn,n}}$ and $\varepsilon_{a_{dyn,n}}$, over the twenty random datasets, averaged over the 100 test cases:

$$\bar{\sigma}_{\theta_{dyn,n}} = \mu_{n_{tests}}[\sigma_{n_{seeds}}[\varepsilon_{\theta_{dyn,n}}]] \text{ and } \bar{\sigma}_{a_{dyn,n}} = \mu_{n_{tests}}[\sigma_{n_{seeds}}[\varepsilon_{a_{dyn,n}}]]. \tag{16}$$

The mean error and error standard deviation for both maximum wave-induced pitch and nacelle acceleration are illustrated in Figure 4. In the Figure, the bars illustrate the mean error, and the error bars the error standard deviation. Overall, the mean error and error standard deviation in the prediction of the structure response show a similar trend to what is observed for the hydrodynamic coefficients. For both maximum pitch and maximum nacelle acceleration, the mean error and error standard deviation decrease sharply up to the database size of 400 samples

and less noticeably afterwards. $a_{dyn,n}$, which is derived from both the structure response in pitch and surge, is usually predicted better than $\theta_{dyn,n}$ by the models in terms of both mean error and error standard deviation. This observation is consistent with the findings on the prediction of hydrodynamic coefficients, where the predicted pitch-related quantities showed the highest mean error and error standard deviation when compared to the ground truth quantities. $\bar{\varepsilon}_{\theta_{dyn,n}}$ is below 17% for all the environmental conditions for the training dataset size of 400 samples, and below 27% at one $\bar{\sigma}_{\theta_{dyn,n}}$ from the mean. For the nacelle acceleration, $\bar{\varepsilon}_{a_{dyn,n}}$ and the error at one standard deviation from the mean are already below 7% and 12%, respectively.

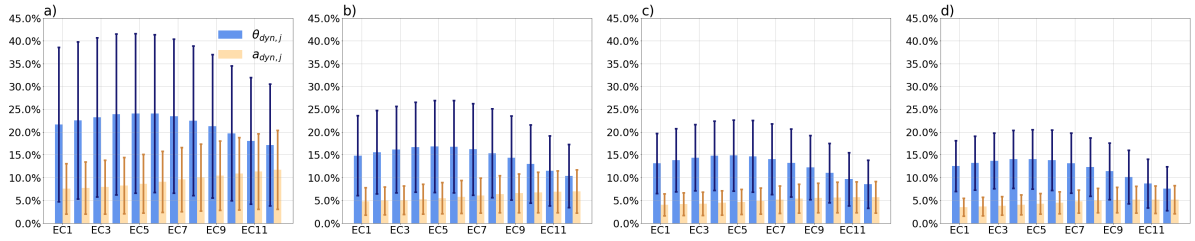

Figure 4: Mean error and error standard deviation on maximum wave-induced pitch response $\theta_{dyn,n}$ and nacelle acceleration $a_{dyn,n}$ prediction with training databases of a) 100 samples, b) 400 samples, c) 800 samples, and d) 1400 samples for all the environmental conditions in Table 4.

5.2. Use case

The optimisation of the semi-submersible is performed for eight different combinations of $w_{C_{se}}$ and $w_{\sigma_{a,w}}$. For each combination of $w_{C_{se}}$ and $w_{\sigma_{a,w}}$, four different training dataset sizes are evaluated. For each dataset size, twenty optimisations are performed, corresponding to an equivalent number of surrogate models. The optimal designs obtained with the surrogate models are fed again into the reference model to evaluate the ground truth value of the weighted standard deviation of nacelle acceleration. For each combination of $w_{C_{se}}$ and $w_{\sigma_{a,w}}$, and each dataset size, the floating substructure capital cost, weighted nacelle acceleration standard deviation, and design features obtained as a result of the optimisation over the twenty models are averaged and compared with the same quantities resulting from carrying out the optimisation with substructure analysis performed with RAFT and HAMS.

Figure 5 a) shows the trade-off between minimising the substructure capital cost and reducing the weighted nacelle acceleration standard deviation, while Figure 5 b) to d) compare the average features of the optimal designs identified with the surrogate model with the ones obtained with the reference model. Only the minimum draft constraint, Equation (4), and the maximum steady pitch constraint, Equation (5), are active for the design that minimises the capital cost of the semi-submersible. This results in the designs obtained with the surrogate models matching exactly the design obtained by performing the optimisation with the reference model. When increasing the weight of the waves-induced nacelle acceleration in the objective function, the designs obtained performing the optimisation with substructure analysis relying on the reference model show a decrease in $R_{out,se}$, and a simultaneous increase in $R_{off,se}$. Reducing the outer column diameter reduces the wave loading on the structure, while the simultaneous increase of the offset between the columns maintains the necessary hydrostatic stiffness to fulfil the maximum steady pitch constraint. The trade-off between the two objective functions is well captured when performing the optimisations with the surrogate model. The design features obtained with the surrogate models at the variation of $w_{C_{se}}$ and $w_{\sigma_{a,w}}$ follow a similar trend as the results predicted with the reference model, although the increase in the semi-submersible draft is largely over-predicted by the surrogate models relying on only 100 training samples for

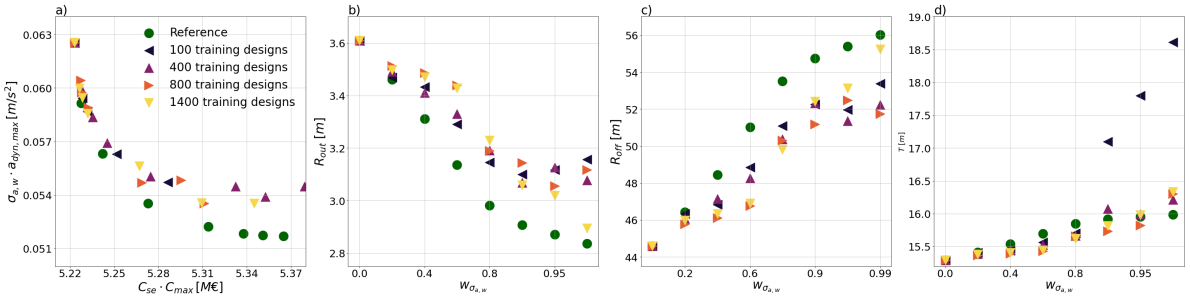

Figure 5: a) Trade-off between minimising the substructure capital cost and reducing the weighted sum of the standard deviation of nacelle acceleration and b), c), and d), average characteristics of the optimal designs identified by RAFT with HAMS (Reference) and RAFT with the surrogate model.

$w_{\sigma_{a,w}} \geq 0.9$. However, evaluating the floating substructure designs with the surrogate model introduces local minima in the objective function, which are not present when evaluating the floating substructure designs with the reference model. These local minima are the result of the variation of the error in the prediction of $a_{dyn,n}$ across the design space and the cause of the difference between the optimal designs identified by the reference model and those identified by the surrogate models.

6. Conclusion

In this work, we introduced a novel surrogate-assisted optimisation approach for the design of floating wind substructures. This method consists of replacing the radiation-diffraction solver necessary to estimate the first order hydrodynamic coefficients for a certain substructure geometry with a data-driven surrogate model, relying on XGBoost tree-based ensemble algorithm. We applied this method to the use case of assessing the trade-off between minimising the capital cost and reducing the wave-induced nacelle acceleration for a semi-submersible supporting a floating wind turbine.

For the semi-submersible geometry considered in this study, the accuracy of the surrogate model in the prediction of the hydrodynamic coefficients increased significantly up to training datasets consisting of 400 designs, while a slighter improvement in accuracy was achieved for larger training datasets. The same trend was observed for the accuracy in predicting the most probable maximum waves-induced pitch response and maximum nacelle acceleration over a 3h storm period. For training dataset consisting of 400 designs, all the hydrodynamic quantities were predicted with a mean error below 7%, whereas the error at one standard deviation from the mean was slightly above 10% only for A_{55}. Overall, pitch-related quantities showed the highest mean error and error standard deviations, while heave-related quantities showed the lowest errors. For the same training dataset size, a lower-than-17% mean error in maximum waves-induced pitch response across all the environmental conditions considered in this study was found when comparing the prediction obtained with the surrogate model with the prediction achieved with the reference model, whereas the error at one standard deviation from the mean was lower than 27%. Concurrently, the mean error in maximum nacelle acceleration in waves was below 7% and the error at one standard deviation from the mean was below 12%.

The results of the optimisation carried out varying the relative weight between capital cost and nacelle acceleration standard deviation showed that the surrogate model captured the trade-off between minimising the floating substructure capital cost and reducing the weighted nacelle acceleration standard deviation and that the designs obtained when performing the optimisation with the surrogate model generally follow the same trend as the optimal designs obtained with

the reference model. However, performing the analysis of the substructure designs with the surrogate model introduced local minima in the objective function that were not present when the analysis was performed with the reference model. These local minima relate to the error in the prediction of nacelle acceleration introduced by analysing the substructure with the surrogate model and caused the discrepancies between the optimal designs individuated by the surrogate model and those achieved by analysing the substructure with the reference model.

The primary reason for developing a data-driven model only for the radiation-diffraction solver instead of the complete frequency domain analysis is to limit the features to floater design parameters. However, this approach does not allow for controlling the error in the response of the floating wind turbine. A data-driven model that directly predicts the response of the floating wind turbine would permit including the error in the loss function during the training phase. This approach will be the focus of future research. The computational cost reduction achieved by using the surrogate model is use case and computational set-up dependent and will generally depend on the number of substructure geometries where radiation-diffraction analysis is avoided. Overall, the surrogate modelling technique presented in this paper would be suited for the preliminary design phase, where it is of interest to identify promising regions of the design space and understand trades and drivers for optimisation of different key performance indicators. Global search methods should be adopted when adopting this surrogate modelling technique to optimisation use cases.

7. Acknowledgments

The project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 860737 (STEP4WIND project, step4wind.eu).

References

[1] Karimi M, Hall M, Buckham B and Crawford C 2017 *J. Ocean Eng. Mar. Energy* **3** 69-97

[2] Hegseth J M, Bachynski E E and Martins J R R A 2022 *Marine Structures* **72**

[3] Leimeister M, Collu M and Kolios A 2022 *Wind Energ. Sci.* **7** 259-281

[4] Ferri G, Marino E, Bruschi N and Borri C 2022 *Renew. Energ.* **182** 1152-1170

[5] Reig M A, Marino E, Mendikoa I, Touzon I and Petuya V 2023 *Ocean Eng.* **285**

[6] Hall M et al. V 2022 *J. Phys.: Conf. Ser.* **2265**

[7] Liu Y V 2021 *Proceedings of the European Wave and Tidal Energy Conference* 2196-1-2196-5

[8] Jonkman J 2010 *Definition of the Floating System for Phase IV of OC3*

[9] Robertson A, Jonkman J and Masciola M 2014 *Definition of the Semisubmersible Floating System for Phase II of OC4*

[10] Chen T, and Guestrin C 2016 *Proc. of the ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining* **22** 785–794

[11] A. Myhr, C. Bjerkseter, A. Agotnes, T. A. Nygaard, 2014 *Renew. Energy* **197**

[12] A. Ioannou, Y. Liang, M. L. Jal´on, F. P. Brennan, 2020 *Ocean Eng.* **66**

[13] G. Rinaldi, A. Garcia-Teruel, H. Jeffrey, P. R. Thies, L. Johanning, 2021 *Appl. Energy* **301**

[14] Baudino Bessone M, Zaaijer M, von Terzi D, Dykes K, Jump E, Viré A 2022 *J. Phys.: Conf. Ser.* **2265** 042018

[15] Mathiesen M, Meyer A, and Kvingedal B 2013 Hywind Buchan Deep Metocean Design Basis, Tech. Report

[16] Gray J S, Hwang J T, Martins J R R A, Moore K T and Naylor B *Struct Multidiscipl Optim* **4** 1075-1104