Coastal Engineering 200 (2025) 104751

Contents lists available at ScienceDirect

Coastal Engineering

journal homepage: www.elsevier.com/locate/coastaleng

Statistical design of submerged artificial oyster reefs using Design of Experiments and clustering strategies

Lei Wang [a], Weikai Tan [b], Marine Thomas [c], Felix Leung [d], Alessandro Stocchino [a,e,*]

[a] Department of Civil and Environmental Engineering, the Hong Kong Polytechnic University, Hong Kong Special Administrative Region of China
[b] Department of Port, Waterway and Coastal Engineering, School of Transportation, Southeast University, Nanjing 210096, China
[c] The Nature Conservancy Hong Kong Foundation Ltd, Hong Kong Special Administrative Region of China
[d] Social Responsibility and Sustainable Development Office, Jockey Club Museum of Climate Change, The Chinese University of Hong Kong, Hong Kong Special Administrative Region of China
[e] State Key Laboratory on Marine Pollution, City University of Hong Kong, Hong Kong Special Administrative Region of China

ARTICLE INFO

Keywords:
Artificial oyster reef
Wave dissipation
Nature-based solutions
Clustering strategies

ABSTRACT

The implementation of artificial oyster reefs as a Nature-based Solution to enhance ecological benefits and shoreline protection represents a prominent area of research. Nevertheless, the wave attenuation performance of multiple underwater artificial reefs has yet to be subjected to comprehensive investigation. To address this gap, we investigated numerically the wave attenuation produced by a sequence of submerged artificial oyster reefs, taking into account a range of incoming wave conditions and configurations of the artificial reefs themselves. A large number of simulations have been designed using an approach based on the Design of Experiment theory, namely the D-optimal approach. The large dataset obtained has been analyzed using unsupervised machine learning techniques, i.e. the weighted K-means. The results showed a clear separation of the combinations of physical variables that led to the lowest transmission coefficients. In particular, three dimensionless variables were identified as being of particular significance for minimizing the transmission coefficient, namely the submergence of the oyster reefs, the length of the oyster reef in relation to the incident wavenumber, and the number of oyster reefs. Relative water depth, wave steepness, distance between adjacent oyster reefs, and seabed slope were found to play a minor role. Based on the results, we suggested an optimal statistical design strategy in order to reach a wave transmission coefficient as low as 0.5, provided the specific characteristic of the site (design wave, slope of the shoaling zone, and water depth). These findings will provide guidance for practical application.

1. Introduction

Oyster reefs, as part of the so-called *ecosystem engineering* (Abascal et al., 2009; Mitsch, 2012), are widely distributed in temperate and sub-tropical estuaries as well as shallow marine areas (Tolley and Volety, 2005; Coen et al., 2007). These reefs contribute significantly to water purification, fish resource enhancement, shoreline erosion mitigation, carbon sequestration, biodiversity enhancement, and overall ecosystem stability (Grabowski and Peterson, 2007; Scyphers et al., 2011; Kellogg et al., 2013; Zu Ermgassen et al., 2013; La Peyre et al., 2014). They play a crucial role in stabilizing intertidal sediments and influencing hydrodynamic patterns within coastal environments. However, due to overharvesting, coastal zone development, water pollution, and disease, most natural oyster reefs face ecological challenges such as habitat destruction, area shrinkage, and functional degradation (Xu et al., 2023). During the past century, more than 85% of oyster reef habitats have been lost (Beck et al., 2011). A striking example can be found in the massive loss of natural oyster reefs recently described in European coastal habitats (Thurstan et al., 2024) as well as South-East Asia (Lau et al., 2020; Williams et al., 2024). Consequently, there is an urgent need to implement practical measures to restore oyster reef habitats, as also demonstrated by the growing interest in policy on local to international scales (Waltham et al., 2020).

Since the 1990s, numerous oyster reef restoration projects have been implemented in estuaries, initially focusing on enhancing the economic benefits of shellfish (Coen and Luckenbach, 2000). Over recent decades, the ecosystem service provided by oysters has been extensively studied (Brown et al., 2014; Graham et al., 2017). These projects often involve constructing mounds from limestone riprap boulders,

* Corresponding author at: Department of Civil and Environmental Engineering, the Hong Kong Polytechnic University, Hong Kong Special Administrative Region of China.
E-mail address: alessandro.stocchino@polyu.edu.hk (A. Stocchino).

https://doi.org/10.1016/j.coastaleng.2025.104751
Received 12 February 2025; Received in revised form 23 March 2025; Accepted 28 March 2025
Available online 9 April 2025

approximately 50 cm in diameter and 1.5–2 meters in height (Pierson and Eggleston, 2014). The establishment of ecosystems such as the artificial oyster reef (EAOR), newly developed artificial oyster reef (NAOR), and non-reef bare substratum (NBS) has demonstrated significant benefits (Wang et al., 2022). The construction of artificial oyster reefs offers considerable advantages for the ecosystem (Goelz et al., 2020) and has emerged as a sustainable solution for soft engineering coastal protection. Large-scale artificial reef installations have been conducted in countries such as for Japan, the United States, South Korea and Hong Kong SAR (Baine, 2001; Hernández et al., 2018; Lau et al., 2020; Williams et al., 2024). These initiatives involve placing structures on the seabed to create suitable habitats for the growth, development, spawning, and shelter of marine organisms (Lee et al., 2018). Oyster reefs, thriving in intertidal zones with their complex structures, are particularly effective at attenuating incident waves, thereby protecting coastlines from erosion (Bridges et al., 2021). This makes them a natural underwater breaker and a valuable additional measure for coastal protection (Morris et al., 2019).

When waves interact with a structure, several processes occur, including wave breaking, reflection, overtopping, and frictional dissipation. For submerged oyster reefs, the effects of reflection and overtopping are reduced compared to emerged reefs where wave attenuation is primarily dominated by these process. The geometric shape of oyster reefs has been shown to resemble that of classical underwater breakwaters (Van der Meer et al., 2005), which enhances wave attenuation as waves pass over these submerged structures. Several dimensionless parameters have been proposed to describe wave attenuation caused by artificial breakwaters, offering empirical relationships based on the interaction between waves and submerged structures (d'Angremond et al., 1996; Seabrook and Hall, 1999; Van der Meer et al., 2005; Buccino and Calabrese, 2007; Srineash and Murali, 2019). The impact on transmission performance has been attributed to two key variables (Srineash and Murali, 2019), the ratio of reef height to water depth (denoted as h_s/d by Seabrook and Hall, 1999) and the ratio of flat reef length to the significant wave height of incident waves (denoted as B/H_i by Van der Meer et al., 2005). As the length of the flat reef increases, the interactions between waves and structures become more complex, and the transmission coefficient decreases as the dimensionless height and length of the flat increase. For instance, when the ratio of reef height to water depth is less than 0.8, the transmission coefficient is greater than 0.6. In restoration project, it has been verified that oyster shell bags exhibit lower wave transmission coefficients (i.e., better wave attenuation effects) due to their compact shape and smaller porosity (Hernández et al., 2018), compared to oyster balls, oyster wave attenuation devices (WADs), and oyster BLK (named, "reef block") (Stokes et al., 2012). The aforementioned studies primarily focus on single submerged reefs. However, for multiple submerged reefs, a special phenomenon known as Bragg resonance (Bragg and Bragg, 1913) must be considered. When specific conditions are met between the arrangement of certain structures and the incident waves, Bragg resonance can cause an abnormal increase in reflected waves, leading to a rapid decrease in transmitted waves (Mei et al., 1988; Bailard et al., 1992; Lin and Liu, 2005; Xie et al., 2011; Liu et al., 2019). This phenomenon is crucial for the optimization process in our study. In the present study we intended to overcome some limitations of previous research, in particular, considering the effect of multiple submerged reefs while expanding the hydrodynamic conditions covering fully nonlinear wave-wave interaction during the wave propagation.

To optimize artificial oyster reefs, it is essential to consider a wide range of potential independent variables, resulting in numerous combinations that need to be calculated. The traditional approach involves sequentially combining different variables, which is time-consuming. Therefore, an effective experimental design that can quickly determine the optimal combinations is crucial. Design of Experiments (DoE) is a powerful tool for data collection and analysis, based on machine learning principles (Telford, 2007). It allows for the simultaneous manipulation of multiple input factors, enabling the determination of their effects on a desired output (response). By manipulating multiple inputs concurrently, DoE can identify significant interactions that might be overlooked when experimenting with one factor at a time. This method can investigate all possible combinations (Full Factorial) (Kabagire and Yahia, 2018; Haddaji et al., 2022) or only a subset of the possible combinations (Fractional Factorial) (Hawkins and Lye, 2006). DoE has been widely applied in academic experimental research and has demonstrated great potential for application (Ba-Abbad et al., 2013; Berger et al., 2017; Bowden et al., 2019; Jankovic et al., 2021). However, its use in ocean engineering has been limited. Notable exceptions include predicting the thrust of a propeller (Islam and Lye, 2009), investigating influence of the environmental conditions on the tension in marine risers for Floating Production Storage and Offloading (FPSO) ships (Hawkins and Lye, 2006), designing a wind turbines and ship structures (Lee et al., 2015), improving the bow shape of a tanker hull (Seok et al., 2019) and evaluating an automatic ocean salt collector (Song et al., 2020).

Additionally, the extensive data processing required for optimizing artificial oyster reefs necessitates the assistance of machine learning techniques. Clustering, a statistical algorithm, can be employed to detect similarities within the dataset (Faber et al., 1994). K-means clustering, an unsupervised learning technique, can be utilized in the data mining process to partition the data into predefined groups, which has been successfully applied in data classification (Koszalka and LaCasce, 2010; Uti et al., 2023; Lagomarsino-Oneto et al., 2024). This method does not require predetermined attributes or the prediction of target values. Unsupervised data mining seeks to uncover hidden relationships and structures within the data (Cawley, 2019). These unsupervised techniques are particularly useful for interpreting raw data without prior knowledge of its structure and for identifying patterns within the data (Chong et al., 2021).

Although significant progress has been made in oyster reef restorations, both academically and in engineering, the optimization of multiple submerged artificial reefs based on the hydrodynamic performance remains unclear. To address this issue, the layout of multiple submerged artificial reefs is numerically optimized using Design of Experiments (DoE) and clustering strategies. The novelty of the present work can be summarized in three key aspects. First, the High-order Spectral method, known for its computational efficiency, is employed to perform parallel simulations using the optimal combinations obtained from the DoE approach, thereby saving considerable time. Second, the weighted K-means method is used to effectively classify the database, which is crucial for subsequent optimization. Third, an optimal strategy is proposed that includes sensitive variables, their ranges, and the corresponding occurrence probabilities of achieving a lower transmission coefficient.

The remainder of the paper is organized as follows. Section 2 introduces the methodology employed in this study, including a description of the fully nonlinear numerical wave model, the Design of Experiments (DoE) approach, data clustering techniques, and the numerical simulation strategy. The results are analyzed in Section 3, followed by a discussion in Section 4. The main conclusions are summarized in the final section.

2. Methodology

2.1. Formulation of the numerical model

In the present study we implemented a numerical wave model based on the High-order Spectral (HOS) method, first proposed by West et al. (1987). HOS models are based on perturbation expansion to higher-order terms of the velocity potential. The Fast Fourier Transform (FFT) employed to solve the velocity potential and its derivative terms

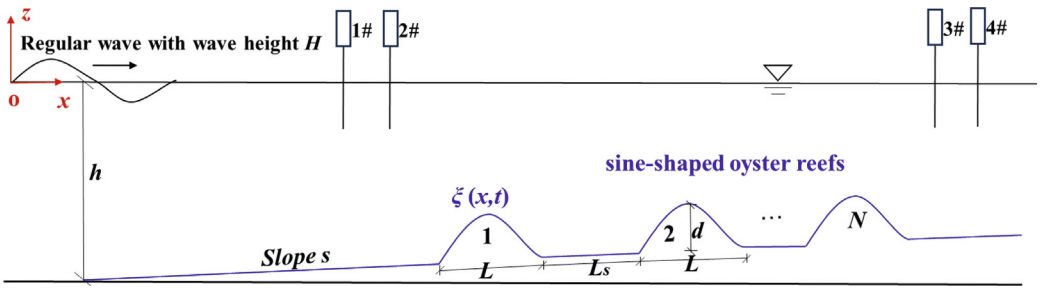

Fig. 1. Sketch of the numerical domain together with an example of artificial oyster reefs. The position of the numerical water level probes is also reported.

significantly enhances computational efficiency. A two-dimensional numerical wave model based on HOS method was formulated to simulate wave propagation over artificial oyster reefs. A sketch of the numerical domain is illustrated in Fig. 1, together with a sketch of the artificial oyster reefs and the relevant variables. In particular, a regular incident wave, with wave height H, is imposed offshore ($x = 0$) and artificial oyster reefs with a sinusoidal shape are uniformly arranged on a mild slope s seafloor. The geometry of the reefs are characterized by their length L, height d, number N, and the space L_s between two successive reefs.

In HOS method, the fluid is assumed to be incompressible, inviscid, and irrotational, based on potential flow theory. A two-dimensional Cartesian coordinate system is defined, with the origin located at the left side of the still water surface plane. The x-axis extends along the direction of the incident wave propagation, while the z-axis is oriented vertically upward, see Fig. 1.

The velocity potential Φ within the domain D, with a working depth of h at the prescribed position (x, z), satisfies the Laplace equation:

$$\nabla \Phi^2 = 0 \qquad \text{in } D \tag{1}$$

where $\nabla = (\partial/\partial x, \partial/\partial z)$.

Hereafter, the mathematical model is solved according to Ducrozet et al. (2012), with the velocity potential split into the sum of a free surface spectral potential component Φ_f and a prescribed additional velocity potential Φ_{add}. The corresponding kinematic and dynamic boundary conditions of the free surface are written as

$$\eta_t = (1 + |\nabla_x \eta|^2)(\Phi_f^s)_z + (\Phi_{add})_z - \nabla_x(\Phi_f^s + \Phi_{add}) \cdot \nabla_x \eta \qquad z = \eta(x,t) \tag{2}$$

$$\begin{aligned}(\Phi_f^s)_t = &-g\eta - \frac{1}{2}(\nabla_x \Phi_f^s)^2 + \frac{1}{2}(1 + \nabla_x \eta^2)[(\Phi_f^s)_z]^2 - (\Phi_{add})_t \\ &- (\nabla_x \Phi_f^s) \cdot (\nabla_x \Phi_{add}) - \frac{1}{(\nabla_x \Phi_{add})^2} - \frac{1}{2}[(\Phi_{add})_z]^2 \qquad z = \eta(x,t)\end{aligned} \tag{3}$$

The inlet boundary condition and the bottom boundary condition can be given as:

$$\frac{\partial \Phi_{add}}{\partial x} = \frac{\omega}{T_r} \eta_{\text{incident}} \qquad x = 0 \tag{4}$$

$$\frac{\partial(\Phi_f^s + \Phi_{add})}{\partial z} - \nabla \xi \cdot \nabla(\Phi_f^s + \Phi_{add}) = \frac{\partial \xi}{\partial t} \qquad z = -h + \xi \tag{5}$$

where ω and η_{incident} are the angular frequency and water surface elevation of the incident waves, respectively, and T_r is the hydrodynamic transfer function, calculated as:

$$T_r = \frac{4 \sinh^2 kh}{2kh + \sinh 2kh} \tag{6}$$

where k is the wavenumber and h water depth

ξ is the topographic transformation. In our case, it is prescribed as Fig. 1

$$\xi = \begin{cases} sx & x \le 0.2L_x \\ sx + d\sin[\frac{\pi}{L}(x - 0.2L_x)] & 0.2L_x < x \le 0.2L_x + L \\ sx & 0.2L_x + L < x \le 0.2L_x + L + L_s \\ \cdots & \\ sx + d\sin[\frac{\pi}{L}(x - 0.2L_x) - (N-1)(L + L_s)] & \\ \qquad 0.2L_x + (N-1)(L + L_s) < x \le 0.2L_x + (N-1)(L + L_s) + L & \\ sx \quad 0.2L_x + (N-1)(L + L_s) + L < x & \end{cases} \tag{7}$$

where L_x is the length of the numerical wave tank.

The initial condition is set for $t = t_0 = 0$ as follows:

$$\begin{cases} \eta(x, t_0) = 0 \\ \Phi_f^s(x, t_0) = 0 \\ \Phi_{add}(x, z, t_0) = 0 \end{cases} \tag{8}$$

The additional velocity potential Φ_{add} can be determined by the incident waves, known as a set of specific basis functions concerning the number of points along the vertical direction of the numerical model. The unknown free surface velocity potential Φ_f^s could be solved as the sum of a series of high-order characteristic functions $\Phi_f^{(m)}$, in which they are decomposed into $\alpha^{(m)}$ and $\beta^{(m)}$ for the consideration of the topographic variation. The former one satisfies the free surface boundary condition and the latter one satisfies the bottom condition, as:

$$\alpha^{(m)}(x,z,t) = \Sigma_{i=0}^{+\infty} \alpha_i^{(m)}(t) \frac{\cosh[|k_i|(z+h)]}{\cosh(|k_i|h)} e^{ik_i x} \tag{9}$$

$$\beta^{(m)}(x,z,t) = \beta_0^{(m)}(t) z + \Sigma_{i=0}^{+\infty} \beta_i^{(m)}(t) \frac{\sinh[|k_i|(z+h)]}{k_i cosh(|k_i|h)} e^{ik_i x} \tag{10}$$

Once the derivative terms of the wave surface elevation $\eta(x,t)$ and the free surface velocity potential $\Phi_f^s(x,t)$ in Eqs. (2) and (3) are calculated, the corresponding variables in the next time step can be obtained using the fourth-order Rung-Kutta method. We refer the reader to Hao et al. (2022) for more details.

To mitigate the influence of reflected waves on numerical accuracy, a damping zone up to 4–5 times the wavelength of the incident waves is implemented at the downstream boundary of the computational domain according to the methodology proposed by Larsen and Dancy (1983). This absorption technique operates through progressive wave energy dissipation within the damping zone, achieved by applying a spatial damping coefficient μ to both free surface elevation and velocity potential during each computational time step. The damping coefficient follows the formulation:

$$\mu = \begin{cases} \exp\left[\left(2^{-d/(n\Delta d)} - 2^{-d_z/(n\Delta d)}\right)\ln \alpha\right] & 0 \le d \le d_z \\ 1 & d_z < d \end{cases} \tag{11}$$

where d represents the distance from the damping zone interior node to the domain boundary, Δd is the grid spacing between adjacent nodes,

Table 1
Range of parameters used for the Design of Experiment.

set	N. of runs	kh	$kH/2$	d/h	$kL/(2\pi)$	$kLs/(2\pi)$	s	N
1	1244	0.5–1	0.012–0.03	0.1–0.7	0.22–0.4	0.11–0.2	0-1:800	1–4
2	11573	0.32–1	0.01–0.03	0.1–0.73	0.18–0.47	0.06–0.25	0-1:800	1–4

d_z is the width of the damping zone, and $n = 5$ and $\alpha = 4$ are empirical constants governing damping intensity and spatial gradient the constants.

Finally, the output of each wave simulation has been used to calculate the transmission coefficient, defined as:

$$K_t = H_t/H_i \tag{12}$$

where H_i denotes the wave height of the incident wave, and H_t denotes the wave height of the transmitted wave. In particular, H_i has been calculated using the two-point method proposed by Goda and Suzuki (1976) based on the time history of surface elevation recorded at two adjacent gauges, namely gauge 1 and gauge 2 (see Fig. 1), within a half-wavelength distance. H_t was computed, with the same method, as the separated incident waves using gauge 3 and gauge 4 as illustrated in Fig. 1.

2.2. Numerical simulation strategy based on D-optimality

To optimize artificial oyster reefs, several potential variables must be considered. In this study, we examined a total of seven variables related to both the hydrodynamic conditions and the geometry of the reefs. Our primary aim is to investigate how the transmission coefficient depends on the combination of these selected variables. To achieve this, we first defined the relevant physical variables and converted them into dimensionless forms. The variables selected include: the relative water depth (kh), wave steepness ($kH/2$), the dimensionless height or submergence of the oyster reef (d/h), the dimensionless length of the oyster reef ($kL/(2\pi)$) and spacing ($kL_s/(2\pi)$), the slope of the seabed (s), and the number of oyster reefs (N). To cover a wide range of variability for these parameters, we set ten different values for kh, $kH/2$, d/h, $kL/(2\pi)$ and $kL_s/(2\pi)$, six values for the bottom slope and varied the number of oyster reefs from 1 to 4. Table 1 provides the detailed values for each variable. Two sets of parameters have been tested to evaluate the influence of the range of chosen parameters on the final results, together with the number of numerical simulations selected using the approach described below.

The values chosen for the relative water depth and wave steepness are representative for realistic coastal conditions (Zhu et al., 2020; Wiberg et al., 2019; Chowdhury et al., 2021; Morris et al., 2021).

Given the variables and their variability, the total number of possible combinations is 2.4 million. Conducting simulations for all these combinations would require an unsustainable amount of computational time. To address this issue and to maximize the likelihood of success in our analysis, a well-designed experiment is recommended to optimize the controlling parameters for achieving the minimum transmission coefficient (Telford, 2007; Kim et al., 2015). The strategy suggested in this study is to implement Design of Experiments (DoE) theories to select the most informative set of experimental conditions. The primary advantage of using the DoE approach is its ability to systematically and efficiently identify the most efficient combinations of factors affecting a process, thereby optimizing outcomes while minimizing the number of required experimental runs and resources.

Compared to linear experimental designs, such as Full Factorial Designs (Kabagire and Yahia, 2018), Fractional Factorial Designs (Hawkins and Lye, 2006), or Response Surface Designs (Lee et al., 2015), alphabetic optimal designs are more suitable for calibrating nonlinear models (Guerrero and Stephen, 2018; Huan et al., 2024) as in the present context of optimizing oyster reef restoration. When generating optimal designs with a focus on precise parameter estimations, D-optimality is the most popular design criterion due to its simplicity in implementation and computation (Singh and Xie, 2020).

For this reasons, we employed the D-optimal criterion to achieve the optimal experimental design in the present study. As in any optimal design, we started defining the parametric statistical model to describe our observations (Huan et al., 2024). Owing to the nonlinearity of process under investigations, we selected a quadratic model for the response y that included all the possible parameter interaction as follows (Myers et al., 2016; Walsh et al., 2024):

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_{ii}^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j + \epsilon \tag{13}$$

where k is the number of experimental parameters (7 in the present case). The model response includes an intercept β_0, k main effects (linear terms), $k(k-1)/2$ two-way interactions and k quadratic terms. The error term, ϵ, is assumed to be independently and identically distributed.

Let p being the number of coefficients of the assumed nonlinear model (36 in this case), any optimal design is based on the construction of the $n \times p$ model matrix $\mathbf{X}$, where n is the number of sets of candidate design points $\boldsymbol{\xi}_n$. The model matrix is then used to compute the symmetric, positive semi-definite $p \times p$ Fischer information matrix as $\mathbf{X}^T\mathbf{X}$ (de Aguiar et al., 1995; Huan et al., 2024). Among the classical alphabetic optimality, D-optimal is formulated in terms of maximization of the determinant of the information matrix, i.e.:

$$max_{\xi_n \in \Xi} det(\mathbf{X}^T\mathbf{X}) \tag{14}$$

where ξ_n is the nth combination of the candidate variables belonging to the set of all possible designs Ξ (Wong, 1994; Cuervo et al., 2016). Maximizing the determinant of the information matrix is equivalent to minimizing the volume of the confidence region of the regression parameters (Meyers et al., 2016; Huan et al., 2024). Compared to other alphabetic optimal designs, such as A, G or E optimal, D-optimal is know to perform better when dealing with nonlinear models prioritizing joint parameter precision, while for example A-optimal favors individual parameter accuracy (Jones et al., 2021; Bodunwa and Adewole, 2022; Iwundu and Israel, 2024). Moreover, the efficiency is straightforwardly computed compared to A and E optimal design and it is generally higher compared to other designs (Wong, 1994; Jones et al., 2021), as we will see below.

D-optimal designs are typically generated using an iterative search algorithm. This algorithm starts with an initial set of experimental points and iteratively adjusts these points to improve the determinant of the information matrix, until convergence. The identified design points minimize the volume of the confidence ellipsoid for the coefficients, thereby determining the optimal combination of potential variables (Melas et al., 2011). The algorithm applied in the present study is the coordinate exchange algorithm implemented in the Matlab function *cordexch* (Cuervo et al., 2016).

For each choice of n, we monitored the behavior of some critical quantities in order to select the final optimal design matrix. In particular, we computed the value of the determinant of the information matrix, the maximum variance function or leverage $d(\xi_i)$ for the ith matrix, the D- and G-efficiency. The definitions of the above quantities read:

$$d(\xi) = \xi_i^T(\mathbf{X}^T\mathbf{X})^{-1}\xi_i \tag{15}$$

$$D\text{-}eff = |\mathbf{X}^T\mathbf{X}|^{1/p}/n \tag{16}$$

$$G\text{-}eff = p/(nd(\xi)) \tag{17}$$

The results of our DOE are shown in Fig. 2 in terms of the efficiencies, leverage, condition number normalized with its maximum value and the log of the determinant of the information matrix. The range of n tested was between 36, the minimum required being $p = 36$, and

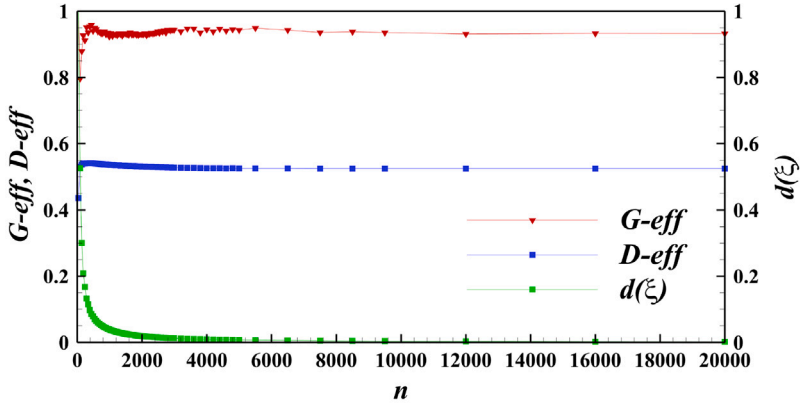

Fig. 2. Evolution of DoE parameters as a function of the number of parameter combinations n: *D-eff*, *G-eff* and leverage $d(\xi)$.

20000. As shown in panel (a), both *D-eff* and *G-eff* rapidly attained their maximum values around $n \backsimeq 400$ and, then, they remained fairly constant. At the same time the variance decayed sharply reaching values as low as 0.039 for $n \backsimeq 1000$.

For the present study, we tested two sets of optimal designs with the aim to study the influence of the initial range of the variables and the influence of the total number of numerical simulations. The two sets, after removing the duplicates, included 1244 and 11573 runs, respectively. For both cases the G- and D-efficiency were very similar (set 1: G-efficiency equal to 0.928, D-efficiency equal to 0.534; set 1: G-efficiency equal to 0.931, D-efficiency equal to 0.524), and with a variance equal to 0.0311 (set 1) and to 0.0031 (set 2).

2.3. Clustering analysis of the simulation output

The main output of the set of simulations performed using the HOS wave model was the computation of the transmission coefficient depending on the combinations of parameters obtained using the D-optimal analysis described in the previous section. The next step was the postprocessing of the data with the aim of identifying a set of parameters that lead to minimum values of the transmission coefficient in the multidimensional parameter space.

To this end, a weighted K-means method (e.g., Modha and Spangler, 2003) is designed and applied to separate the HOS dataset into several clusters mainly according to transmission coefficients, i.e., different clusters should have rather distinct values of transmission coefficients. We selected the K-means method primarily because it is a weight-sensitive clustering approach, allowing us to separate data into distinct clusters by adjusting the weights of variables. This property ensures that each cluster can be characterized by distinct K_t values, which is crucial for our analysis. Unlike some other clustering methods, such as hierarchical clustering or Gaussian mixture models, K-means offers the flexibility to emphasize certain variables over others through weighting, thereby enhancing the separation between clusters. Indeed, such a strategy is feasible as Ackerman et al. (2021) mathematically showed that the K-means method is weight sensitive, i.e., it can separate data into any set by altering the weights of variables. In the discussion section, we will show that the weighted K-means method groups data with different transmission coefficients. Afterward, the distribution of 7 input variables within each cluster will be investigated to determine the key parameters that dominate the wave attenuation.

A brief introduction of the weighted K-means is given as follows. Before clustering, an appropriate number of clusters (K) is selected. Given a set of K, the weighted K-means algorithm is repeatedly executed and a sum of squared errors (SSE, summed distance of each sample to its own centroid), is calculated. The K at the inflection point of the $K - SSE$ curve is considered as the number of clusters.

Consequently, given initial weights for 8 variables (7 input variables and transmission coefficient) and randomized cluster centroids, the target function (cf., Huang et al., 2005)

$$J_w = \Sigma_{n=1}^{N} \Sigma_{k=1}^{K} \Sigma_{j=1}^{M} \delta_{nk} \omega_j Dist(x_{n,j}, \mu_{k,j}) \tag{18}$$

is minimized by adjusting δ_{nk}, where δ_{nk} is 1 when the sample n belongs to cluster k, and 0 otherwise. N is the number of samples, $M = 8$ is number of variables, ω_j is the weight of variable j, $x_{n,j}$ is the variable j of sample n, $\mu_{k,j}$ is the variable j of cluster k centroid. Here, $Dist$ is the function to calculate distance, which is simply taken as Euclidean distance. Once δ_{nk} is determined through minimizing J_w, the centroid of each cluster is updated through,

$$\mu_k = \frac{\Sigma_n \delta_{nk} x_n}{\Sigma_n \delta_{nk}} \tag{19}$$

Iterations are then conducted until convergence.

The weights ω_j for each variable were selected by ensuring each cluster had rather distinct transmission coefficients. A sensitivity analysis will be presented in the Results section.

Finally, the Sum of Squared Errors (SSE) applied in the K-means method can measure the degree of the cluster coherence. The SSE was calculated as follows:

$$SSE = \sum_{i=1}^{k} \sum_{p \in C_i} |p - m_i|^2 \tag{20}$$

where m denotes the centroid of the cluster C_i within the samples p, and k denotes the number of the cluster. The smaller the SSE, the closer the data points are to the corresponding center, and the better the clustering effect. The values of the SSE as a function of the cluster numbers was used for selecting the optimal number of clusters.

3. Results

3.1. Wave model validation

The wave model based on High-order Spectral (HOS) method was initially validated against available experimental observations. In particular, we used laboratory measurements conducted by Beji and Battjes (1993), where regular waves climbed a mild slope and passed through the submerged breakwater. The laboratory setup for this experiment is illustrated in Fig. 3.

To quantitatively compare the numerical model results and the experimental observations, we computed several statistical indexes. In particular, we evaluated the correlation coefficient (CC), the $BIAS$, the root mean square error ($RMSE$), the Hanna and Heinold indicator (HH) and the model skill score ($SKILL$), as follows:

$$CC = \frac{\sum (S_i - \bar{S}) \cdot (O_i - \bar{O})}{\sigma_S \cdot \sigma_O \cdot N} \tag{21}$$

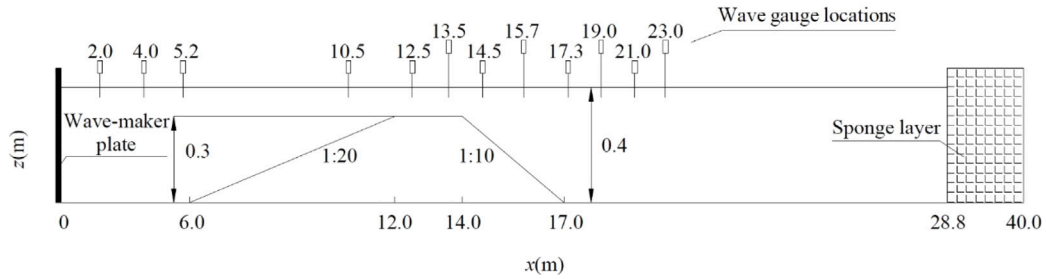

Fig. 3. Sketch of the geometry used during the experiments (Beji and Battjes, 1993). The position of the wave gauges is also reported.

Table 2
Statistical parameters.

	CC	*BIAS*	*RMSE*	*HH*	*SKILL*
λ/dx	spatial discretization				
37	0.992	0.0037	0.1146	0.991	0.980
56	0.992	0.0007	0.1133	0.991	0.981
75	0.991	−0.0019	0.1153	0.990	0.980
dt	integration time				
0.04	0.976	−0.0061	0.1833	0.975	0.949
0.02	0.987	−0.0035	0.1379	0.986	0.971
0.01	0.991	−0.0019	0.1153	0.990	0.980
M	nonlinearity				
10	0.991	−0.0065	0.1197	0.990	0.978
12	0.991	−0.0019	0.1153	0.990	0.980
14	0.991	−0.0012	0.1173	0.990	0.979

$$BIAS = \frac{\sum(S_i - O_i)}{N} \tag{22}$$

$$RMSE = \sqrt{\frac{\sum(O_i - S_i)^2}{N}} \tag{23}$$

$$HH = \sqrt{\frac{\sum(S_i - O_i)^2}{\sum(S_i \cdot O_i)}} \tag{24}$$

$$SKILL = \frac{\sum(O_i - S_i)^2}{\sum(O_i - \bar{O})^2} \tag{25}$$

where N is the total number of samples, S_i and O_i are the simulated and observed variables with their respective means $\bar{S}$, $\bar{O}$ and standard deviations σ_S, σ_O. HH is also referred as a symmetrically normalized root mean square error: it takes into account both the average and scatter components of the error and it is unbiased towards simulations that underestimate the average.

The time series of the free surface recorded at the station $x = 14.5$ m has been selected for comparison with the numerical results and to perform convergence tests. The results of the sensitivity analysis is shown in Fig. 4 and the corresponding statistical indexes are reported in Table 2.

In particular, we verified the convergence against the spatial discretization, i.e. varying the number of grids per wavelength (λ/dx), varying the integration time step, and the order of nonlinearity, panels (a), (b) and (c), respectively. For the range of parameters investigated, the correlation coefficient is consistently around 0.9 or above, as the $SKILL$ score and the Hanna and Heinold indicator (HH). The results regarding the order of nonlinearity show that the difference of the statistical indicators are very small for the three values tested. Based on this analysis, we selected the following set: $\lambda/dx = 56$, $dt = 0.02$ s and $M = 12$.

With the selected numerical parameters, we performed simulated several regular waves simulations and compared with the experimental observations. The comparison of water surface elevation time series between numerical results (represented by red dotted lines) and experimental data (represented by black solid lines) is shown in Fig. 5 for four measuring stations ($x = 5.2, 12.5, 13.5$ and 19 m, see Fig. 3 for their location). The observed agreement between the experimental measurements and the numerical simulation is very good either in terms of wave height and phase, except for the initial stage of the measurements while the wave was developing. The statistical indexes corresponding to the test shown in Fig. 5 are reported in Table 3.

Table 3
Statistical parameters of the comparison of the simulated results against the experimental observations in multiple locations. Note that for computing the statistics the normalized free surface elevation η/A has been used.

x (m)	*CC*	*BIAS*	*RMSE*	*HH*	*SKILL*
5.2	0.996	−0.000108	0.0644	0.996	0.992
12.5	0.985	−0.000807	0.1593	0.985	0.971
13.5	0.991	−0.008389	0.1185	0.991	0.982
19	0.995	−0.037545	0.1137	0.991	0.979

Thus, the present HOS model proved to be able to accurately reproduce wave propagation over submerged structures.

3.2. Data clustering results

3.2.1. Optimal number of clusters

The vanilla K-means method needs a pre-defined number of clusters (K). To select K, we follow the *elbow* of inertia method to find the appropriate K value (Cremonini et al., 2021; Tan et al., 2024; Daliri et al., 2025). In this method, the sum of squared errors (SSE) is calculated, see Eq. (20), for a set of given K. SSE measures the samples' total distance to their own centroid after data clustering. When SSE is large, the clustering yields unreliable results as samples are on average far from their centroids. Suppose a small value of SSE is obtained, the clustering results are considered as acceptable. The SSE value usually drops quickly with increasing K, then descends slowly as K further increases. Therefore, for the balance between efficiency and accuracy, the inflection point of SSE curve (against K), the so-called *elbow* is commonly chosen as the best number of clusters for subsequent analysis.

Fig. 6 shows the SSE as function of cluster number K obtained starting from the output of numerical simulations belonging to set 1. Note that both set 1 and set 2 lead to almost identical results. The SSE decreases rapidly at first and then stabilizes as the value of K increases, indicating the convergence of error accumulation. To ensure that the optimal configuration is as precise as possible, it is important to consider both error convergence and sufficient classification. Based on the trend of the SSE we selected for the present analysis a number of clusters equal to 6.

3.2.2. Selection of the variables weights

After determining the optimal classification strategy in terms of number of clusters, we tested four groups of weights for the seven variables. The values of the weights are shown in Table 4. The weights

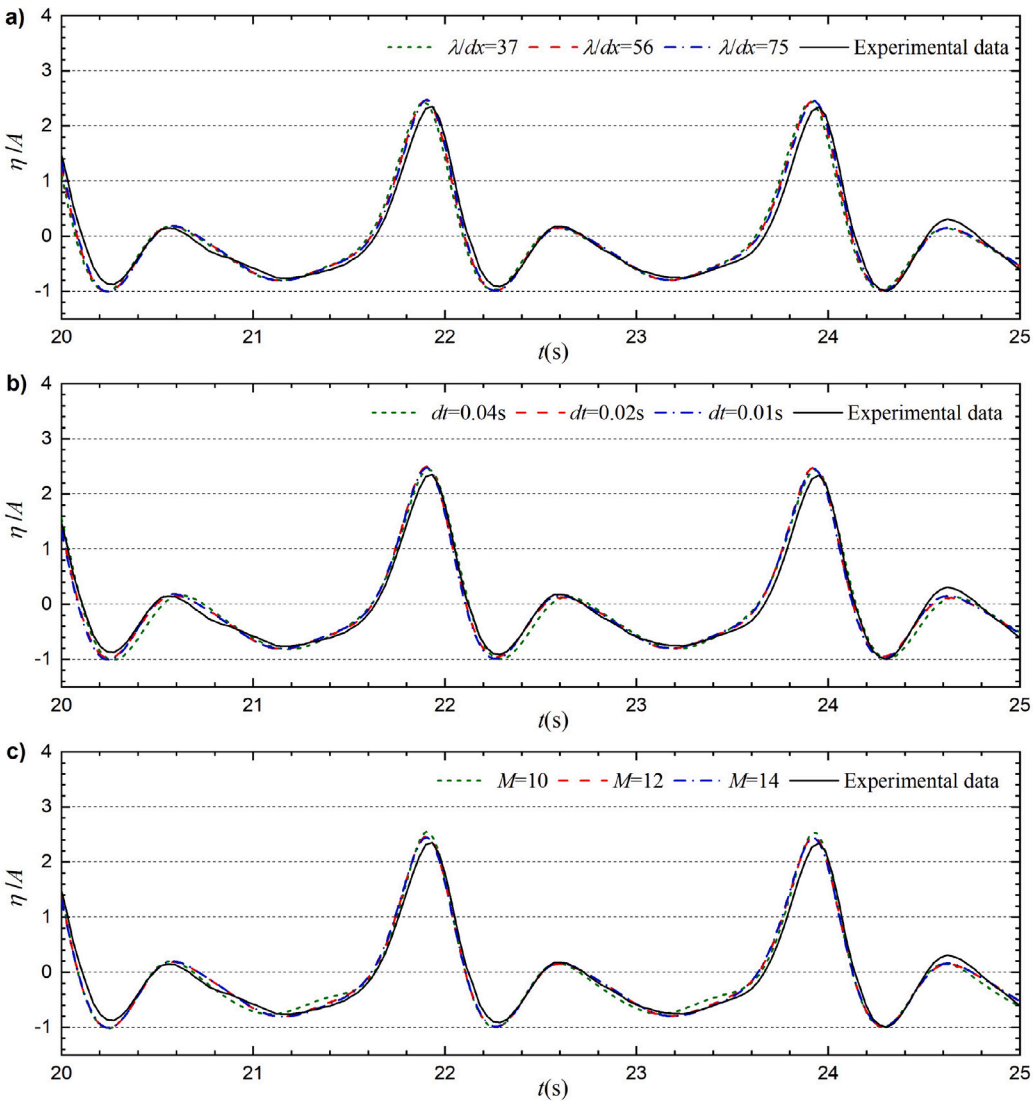

Fig. 4. Convergence analysis of the numerical model: (a) effect of spatial discretization; (b) effect of decreasing integration time;(c) effect of increased nonlinear order. In each panel the experimental free surface time signal at x = 14.5 m is reported for comparison.

Table 4
Weights used in the K-means algorithm.

kh	$kH/2$	d/h	$kL/(2\pi)$	$kLs/(2\pi)$	s	N	K_t
0.0857	0.012–0.03	0.0857	0.0857	0.0857	0.0857	0.0857	0.5
0.0571	0.0571	0.0571	0.0571	0.0571	0.0571	0.0571	0.6
0.0286	0.0286	0.0286	0.0286	0.0286	0.0286	0.0286	0.8
0.0143	0.0143	0.0143	0.0143	0.0143	0.0143	0.0143	0.9

have been arbitrarily decided imposing an increasing value to our target variable K_t and, then, assigning an equal weight to the other variables such that the sum of the weights was unitary.

The weights assigned to the variables were carefully selected to ensure that each cluster exhibits distinctly different K_t values. We progressively increased the weights of K_t until the K_t values were well-separated among the clusters.

We evaluated the performance of the clustering analysis on the two sets of numerical simulations varying the four groups of weights. To better identify the best choice of the weights, we adopted a dimension reduction method named T-distributed Stochastic Neighbor Embedding (T-SNE), which projects the clustered results from high-dimensional variable space to 2-dimensional latent space (Hinton and Roweis, 2002). This approach is based on the definition of two probability distributions in a high-dimensional variable space for the original data, in the form of a Gaussian neighborhood distribution, and a 2-dimensional latent space (for data visualization), again a Gaussian neighborhood distribution with unitary variance. As the data projection from high-dimensional space to the 2-dimensional latent space is not known, a training process is designed by reducing the difference between the two distributions, computing the sum of Kullback–Leibler divergences of the two distributions as a measure of their differences.

The result of the reduction from a multidimensional space to a two-dimensional one is shown in Fig. 7 for all groups of weights ω_j. Six clusters are represented by different symbols and colors. There is almost no overlap among the points belonging to different clusters suggesting that the clustering procedure efficiently classified and separated the multi-dimensional dataset.

From panel (a) to (d) the weight of the transmission coefficient increases from 0.5 to 0.9. Values below 0.8 lead to a poorer distinction among the six clusters, see panel (a) and (b) where the points of different colors are mixed together especially for clusters 3, 4, 5 and 6. When the weight for K_T is equal to 0.8 (panel c) the six clusters are very well distinct. A further increase in the weight does not produce significant improvements.

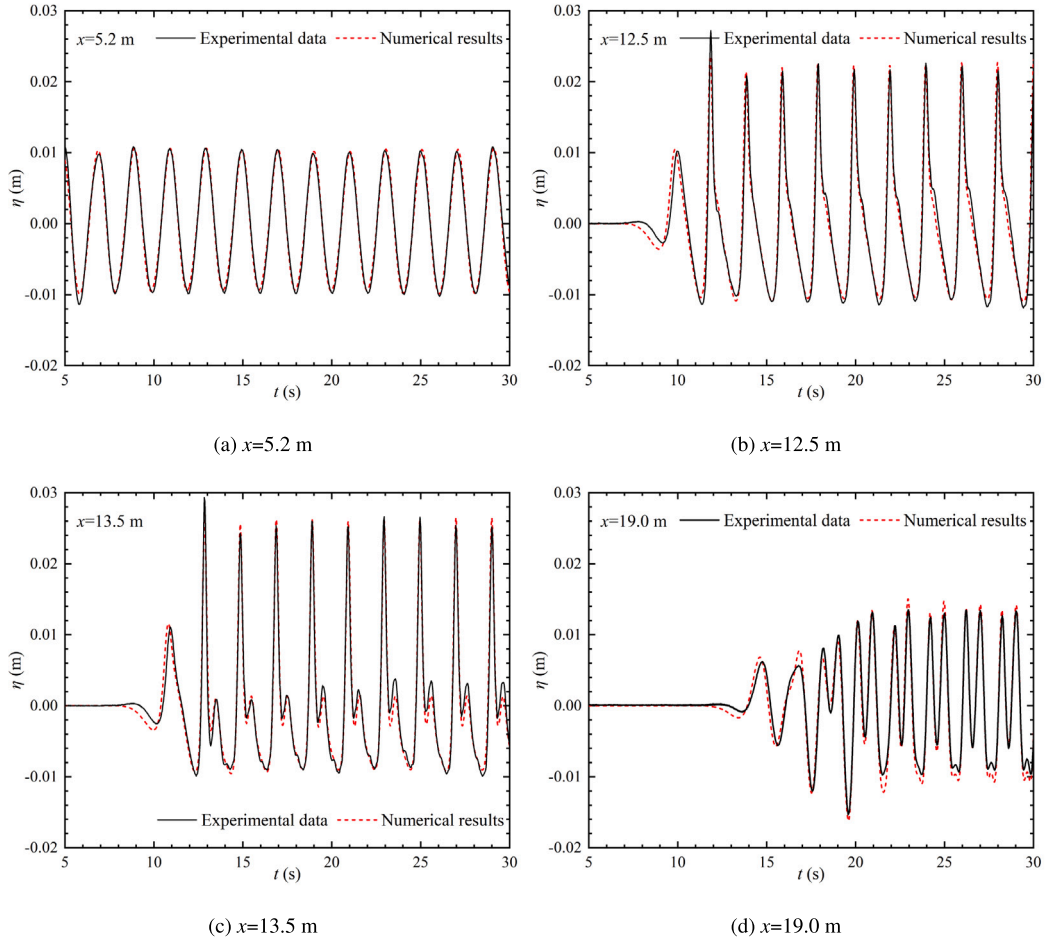

Fig. 5. Comparison of the time signal of water surface elevation between the numerical results and experimental data at four locations, see Fig. 3.

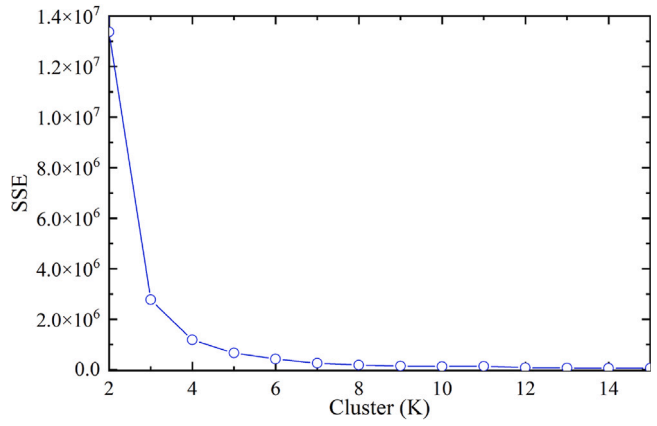

Fig. 6. Sum of Squared Errors (SSE) as a function of the number of clusters K.

This conclusion is even clearer if the actual distribution of K_t values corresponding to the different clusters is analyzed, see Fig. 8. The clustering results are arranged stepwise according to the transmission coefficient, from smallest to largest corresponding to Cluster 1 to Cluster 6. The vertical axis reports the probability of occurrence for each cluster. The same weight for K_t that led to a clear distinction of the clusters in the T-SNE visualization results in non overlapping values for our target variable. Based on this analysis, a weight of 0.8 for K_T has been selected and the results shown in the rest of the paper will refer to this choice.

3.2.3. *Influence of the number of simulations and parameter range*

As discussed in Section 2.2 two sets of simulations have been identified based on the D-optimal design. Here we show the results of the clustering analysis on the two sets for the purpose of discussing the potential role of the number of simulations and the range of the main variables.

The distributions of K_t among the six clusters is shown in Fig. 9, panel (a) corresponding to set 1 (smaller number of simulations) and panel (b) corresponding to set 2 (larger number of simulations). As expected for both cases, the weighted K-means produced a well separated clustering leading to non overlapping values of the transmission coefficient. In light of searching for the best configuration in terms of minimizing the transmission coefficient it could be argued that the increase of number of simulations does not change substantially the final output. Note that for both cases the G- and D-efficiency were already almost constant and the decrease of the variance was relatively small from set 1 and set 2. This implies that once the D-optimal set of parameters has been select, increasing the number of simulations does not produce important differences.

Moreover, the results suggest that does exist a combination of parameters that lead to an efficient wave attenuation of the submerged oyster reefs. In fact, Cluster 1, for both sets of simulations, is composed by combinations of variables that yield to K_t between 0.19 and 0.51 and between 0.1 and 0.38 for set 1 and 2 respectively. Similar values have been recently observed in laboratory studies of wave attenuation

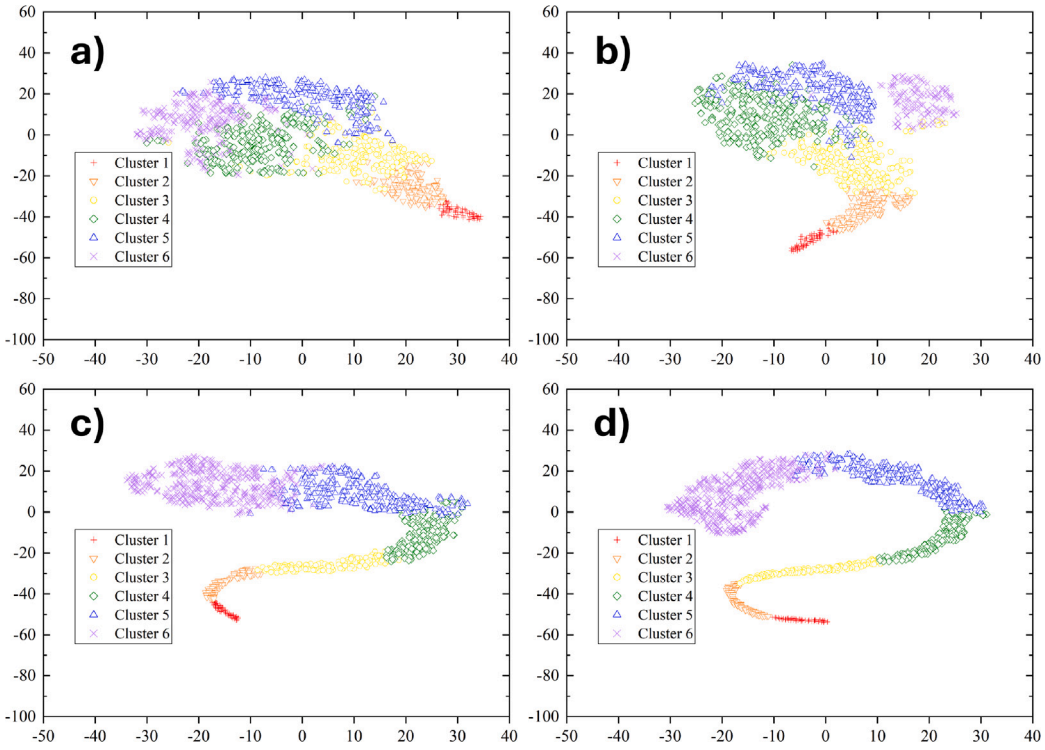

Fig. 7. T-SNE visualization of the clustered results obtained varying the variables weights ω_j, each colored symbol corresponds to a single cluster. (a) K_t weight equal to 0.5; (b) K_t weight equal to 0.6; (c) K_t weight equal to 0.8; (a) K_t weight equal to 0.9.

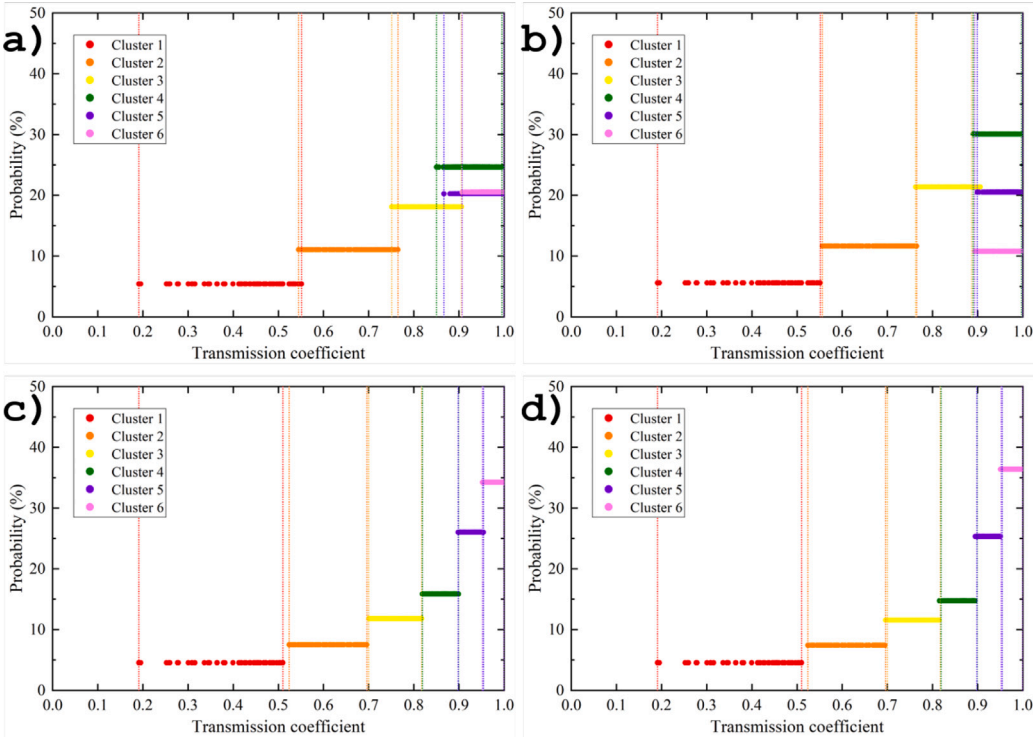

Fig. 8. Distribution of the transmission coefficient and its probability corresponding to the six clusters. (a) K_t weight equal to 0.5; (b) K_t weight equal to 0.6; (c) K_t weight equal to 0.8; (a) K_t weight equal to 0.9.

with submerged artificial reefs or breakwater with impermeable and porous conditions (van Gent et al., 2023; Huang et al., 2024; Provan et al., 2024; Huang et al., 2025). The values of K_t remains reasonably small (i.e. below 0.7) also in Cluster 2 for set 1 and Cluster 2 and 3 for set 2, whereas the other clusters show higher values, suggesting that, for the corresponding multidimensional parameters, the design of the artificial reef do not produce efficient wave attenuation. The detailed range of values for both set 1 and set 2 is reported in Table 5.

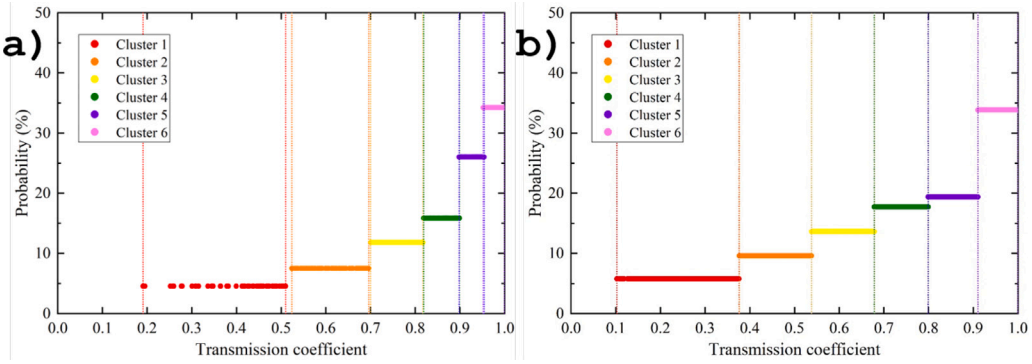

Fig. 9. Distribution of the transmission coefficient and its probability corresponding to the six clusters. (a) set 1; (b) set 2.

Table 5
Range of variability of the transmission coefficient K_t for each cluster for set 1 and 2.

Cluster N.	K_t set 1	K_t set 2
1	[0.19, 0.52)	[0.10, 0.38)
2	[0.52, 0.70]	[0.38, 0.54]
3	[0.70, 0.82]	[0.54, 0.69]
4	[0.82, 0.90]	[0.69, 0.80]
5	[0.90, 0.95]	[0.80, 0.91]
6	[0.95, 1.00]	[0.91, 1.00]

Considering the probability of each cluster, it is worth noting the most efficient combination of parameters included in Cluster 1 has the lowest probability, around 5%. This implies that the optimal design conditions, at least in the range of parameters investigates, are the least probable among within the whole parameter space. This aspect will be further investigated in the next sections. In the following we will consider the results obtained for the smaller set of simulations since it already provide clear information on the performances in terms of wave attenuation. The results for the larger data set are reported in the Supplementary Information document.

4. Discussion

4.1. Influence of single independent variables on the transmission coefficients

In this section we show the sensitivity analysis considering the influence of each independent variables separately on K_t. Box plots are used to describe the performance of wave transmission influenced by the selected seven variables. The data will be arranged in terms interquartile range (IQR). In particular, 50% of the data will be represented a rectangular box bounded by first quartile (Q_1), lower boundary of the box, and the third quartile (Q_3), upper boundary of the rectangle box. Moreover, the statistical median or the second quartile (Q_2) is represented as a straight line within the box and the mean value is shown with a square symbol. Points that fall outside the rectangle are considered outliers, with the maximum at the distributional range of $Q_3 + 1.5IQR$ and the minimum at the distributional range of $Q_1 - 1.5IQR$.

We first consider the influence of the hydrodynamic conditions and the domain geometry on the transmission coefficient. Fig. 10 shows the distribution of K_t as a function of the relative wave depth (kh), panel (a), the wave steepness ($kH/2$), panel (b), and the bottom slope (s), panel (c). Generally, the transmission coefficient is found relatively high with a median value ranging between 0.7 and 0.9 and a mean values varying between 0.65 and 0.8 in the range investigated for the three variables considered. A few outliers appear in the lower coefficient values range. On the contrary, a much more pronounced variability is observed when the geometry of the oyster reefs is considered. Fig. 10 shows the distribution of K_t as a function of the relative submergence (d/h), panel (d), the dimensionless reef length ($kL/(2\pi)$), panel (e), the dimensionless reef spacing ($kL_s/(2\pi)$), panel (f), and the reef number (N), panel (g).

As far as the relative submergence is concerned, for smaller relative submergence (d/h), the transmission coefficient is concentrated around 0.95 with a very narrow distribution. As d/h increases, the distribution of the K_t broadens, reaching much lower values. When d/h reaches its maximum value, the minimum transmission coefficient can be as low as 0.10. This suggest, as expected, that a larger ratio of reef height to local water depth enhances the wave attenuation.

Regarding the effect of different lengths of artificial reefs, panel (e), shorter relative lengths appear more effective in attenuating the incoming waves, as shown by the lower values of the transmission coefficients. This might be due to the fact that smaller relative lengths of oyster reefs correspond to longer incoming waves, which tend to be more easily transmitted over submerged artificial reefs. As ($kL/(2\pi)$) increases, K_t distributions tends to be narrower and centered on values around 0.9.

The spacing between consecutive reefs relative to the wavelength seems to slightly affect the wave transmission, panel (f), leading to distribution centered around $K_t \approx 0.9$, at least in the range of values investigated.

Finally, the influence of the number of reefs on K_t is shown in panel (g). The trend is quite straightforward to explain: the more oyster reefs there are, the more incident wave energy is attenuated, resulting in reduced wave transmission.

4.2. Bivariate analysis

Using the K-means approach, data with similar attributes are classified together. However, it is important to identify what these similar attributes are and how the independent variables are distributed across different clusters. We use bivariate analysis, which combines a scatter plot and frequency distributions, to identify the effect of the main variables on our target function. Fig. 11 illustrates the distributions of independent and dependent variables in each cluster dividing, as in Fig. 10, the set of variables linked to the wave conditions and reef geometries. Different clusters are represented by different colors ranging from Cluster 1 to Cluster 6. The scatter plot shows the transmission coefficients along each independent variable, with the frequency distribution of the independent variable displayed at the top and that of the transmission coefficient on the right. The statistical parameters of the frequency distributions (mean, standard deviation and skewness) for each variable are reported in Table 6.

In Fig. 11, several independent variables can be observed evenly distributed among different clusters, without no distinct dominance, such as relative water depth in panel (a), wave steepness in panel (b), and slope of the seabed in panel (c). This suggests that the conditions

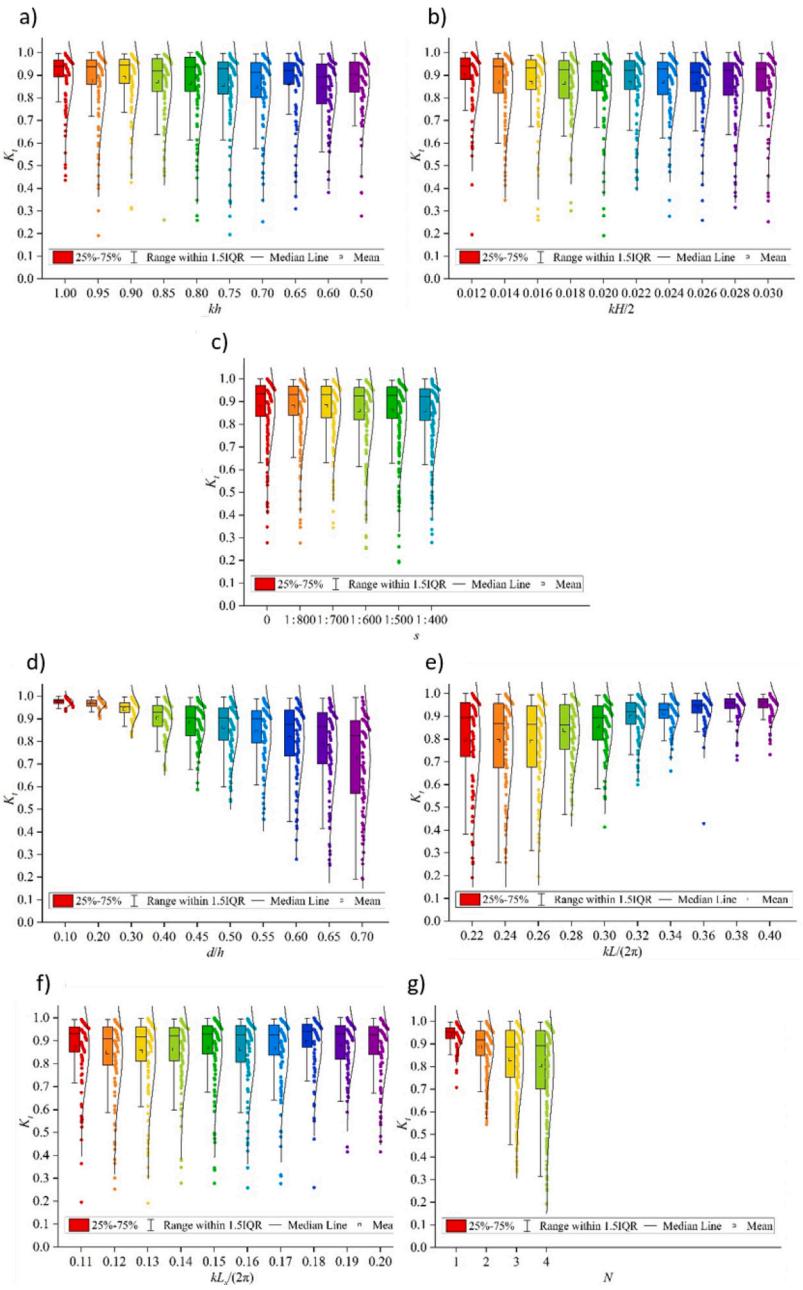

Fig. 10. Box plots of the transmission coefficient (K_t) as a function of: (a) kh; (b) $kH/2$; (c) s; (d) d/h; (e) $kL/(2\pi)$; (f) $kL_s/(2\pi)$ and (g) N.

of incoming waves and the seabed slope has a minor influence on the wave attenuation performance. Xie (2022) pointed out that a sloping seabed can shift the resonance frequency, but has little effect on wave transmission, consistent with our observation.

On the other hand, the distribution three of the other independent variables plays a dominant role on controlling the wave transmission as shown in Fig. 11, namely the relative submergence (panel (d)), the length of the oyster reef (panel (e)), and the number of oyster reefs (panel (g)). The distance between successive reefs is again found to slightly influence the wave transmission performance, see panel (f).

These observations are in line with previous studies on submerged breakwaters (Bredes et al., 2022) and less traditional submerged structures (van Gent et al., 2023; Huang et al., 2024; Provan et al., 2024; Huang et al., 2025).

Regarding the role of the relative submergence (d/h), Fig. 11 panel (d), it is important to note how the shape of the relative frequency distribution within the different clusters varies from Cluster 1 to 6. The former is characterized by a distribution with a small standard deviation and a negative skewness, whereas the distributions for the other clusters have much higher standard deviation and comparable skewness, see Table 6. In Cluster 1, the narrower range of relative submergence (i.e., $d/h \in [0.55, 0.70]$) results in a relatively narrow distribution in K_t around a mean value of about 0.4, i.e. only 40% of the wave is transmitted. In contrast, the wider range in Cluster 6 (i.e., $d/h \in [0.10, 0.70]$) allows more than 97% of waves to pass over the reef. This observation is in contrast with the results discussed in Morris et al. (2021), where it was suggested the for dimensionless height of the reef less than 0.8, the transmission coefficient is greater than 0.6. The reason for this discrepancy can be found in the different reef geometry (shape and number of reefs). Other studies showed that modifying the reef layout could reduce the wave transmission (Chang and Liou, 2007; Zhang et al., 2012; Peduzzi et al., 2012).

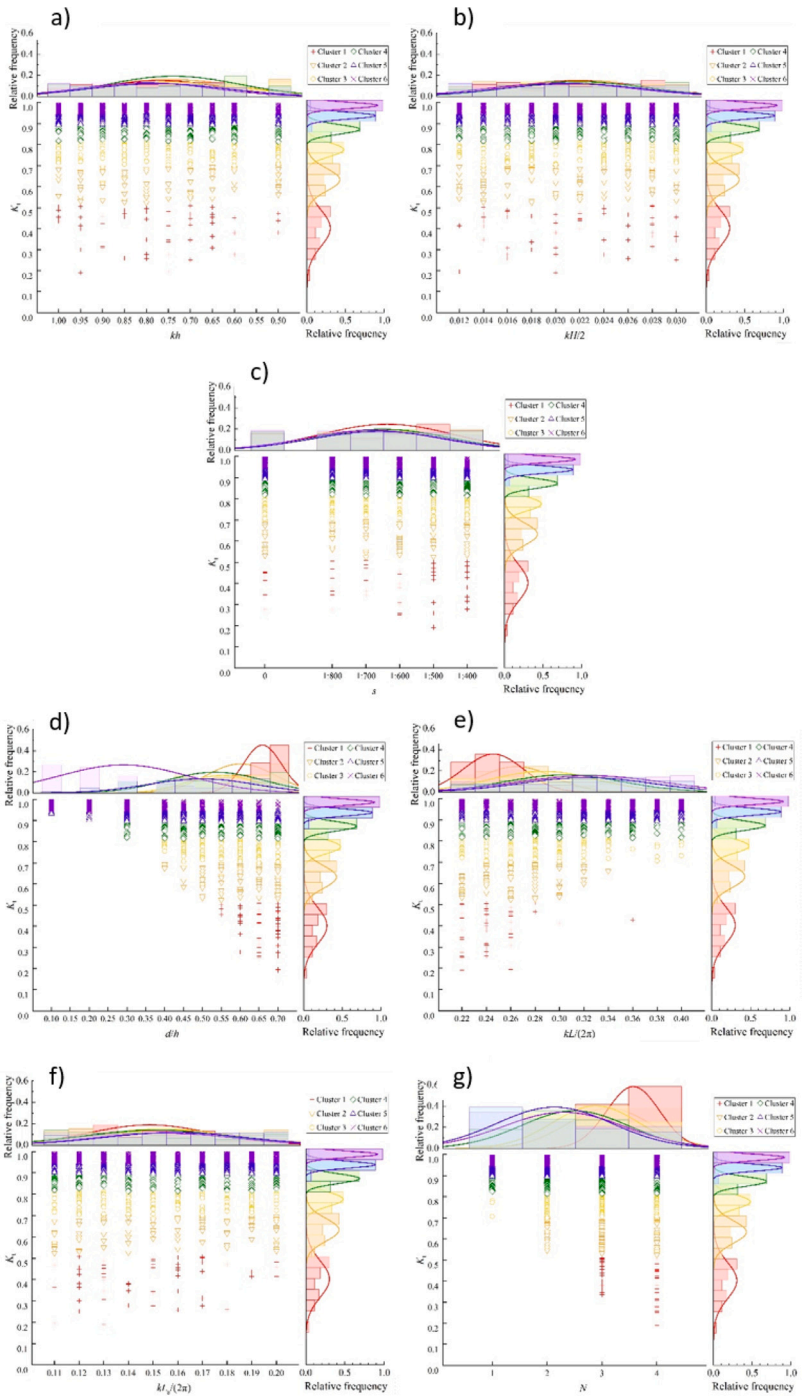

Fig. 11. Bivariate plots and corresponding frequency distributions of the transmission coefficient (K_t) as a function of: (a) kh; (b) $kH/2$; (c) s; (d) d/h; (e) $kL/(2\pi)$; (f) $kL_s/(2\pi)$ and (g) N.

As far as the relative length of the oyster reef is concerned, a similar trend can be observed, see Fig. 11 panel (e). From Cluster 1 to Cluster 6, the probability distribution of the independent variable tend to produce higher mean values and higher standard deviations. Unlike the trend observed in Cluster 1, where the probability distribution of the independent variable increases with relative submergence, there is no significant change in the relative length of the oyster reef.

Regarding the number of underwater artificial reefs, it was shown that it affects wave transmission (Wang et al., 2006; Tsai et al., 2011). However, no quantitative indications were suggested. As shown in Fig. 11 panel (g), a lower transmission coefficient is observed when the number of oyster reefs is greater than two. Furthermore, as the number of oyster reefs increases, the relative frequency peak increases and the transmission coefficient value decreases accordingly. The present results suggest that the optimal dimensionless reef length ($kL/(2\pi)$) in Cluster 1 is around 0.5, leading to low K_t values. The reduction in K_t as a function of the submerged reef length could be attributed to the well-know Bragg resonance effect (Bragg and Bragg, 1913). Indeed, when the length of the periodic perturbation (reef) of the seabed is close to half of the incident wavelength, Bragg resonance occurs as observed in laboratory experiments (Li et al., 2020; Gao et al., 2024) and demonstrated in theoretical analysis (Chang and Liou, 2007; Liu et al., 2020; Xie, 2022). The resonance effect can result in an increase in reflected waves and a decrease in transmitted waves.

Table 6
Statistical parameters of the frequency distributions shown in the bivarate plots of Fig. 11.

		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
K_t	Mean	0.3986	0.6227	0.7678	0.8593	0.9238	0.9743
	STD	0.0877	0.053	0.0354	0.0232	0.0166	0.0059
	skewness	−0.6128	−0.4869	0.2132	−0.8018	−0.4296	−8.1659
kh	Mean	0.7708	0.7494	0.7534	0.7407	0.7784	0.8035
	STD	0.1358	0.156	0.1589	0.1474	0.1514	0.1436
	skewness	−0.0062	−0.1868	−0.0424	0.0788	−0.1751	−0.3894
$kH/2$	Mean	0.0218	0.0204	0.0218	0.0218	0.0212	0.0203
	STD	0.0056	0.0056	0.0055	0.0055	0.0058	0.0058
	skewness	−0.0644	0.1305	−0.1579	−0.08	−0.1272	0.148
s	Mean	488	497	507	496	492	496
	STD	243	251	256	250	258	270
	skewness	−0.829	−0.9156	−0.9128	−0.8814	−0.8392	0.8268
d/h	Mean	0.6547	0.5977	0.5496	0.5302	0.4948	0.2892
	STD	0.0498	0.0903	0.0934	0.1286	0.1418	0.1706
	skewness	−0.7632	−0.5518	−0.0169	−0.2376	−0.5561	0.6083
$kL/(2\pi)$	Mean	0.2449	0.2639	0.2861	0.3029	0.3211	0.324
	STD	0.0245	0.03	0.0444	0.0483	0.054	0.0599
	skewness	2.0369	0.2169	0.4851	0.0549	−0.3162	−0.3678
$kL_S/(2\pi)$	Mean	0.1481	0.1554	0.1533	0.1503	0.1583	0.154
	STD	0.0247	0.0304	0.0277	0.0293	0.0287	0.0285
	skewness	0.3912	−0.0037	0.0952	0.238	−0.1618	0.0026
N	Mean	3.5849	3.1264	2.8788	2.4826	2.1007	2.2768
	STD	0.4927	0.785	0.8074	1.0028	1.0822	1.1375
	skewness	−0.3446	−0.226	−0.0355	0.0998	0.5145	0.2808

4.3. Optimal strategy

The final question to be answered is whether meeting all these conditions ensures the best wave attenuation performance, or what is the probability of achieving the lowest transmission coefficient. To this end, we consider the combined effect of the three most variables, namely d/h, $kL/(2\pi)$ and N.

Fig. 12 shows the bivariate plot of the combined effect of the relative submergence and length of oyster reefs. The probability distribution of relative submergence is displayed at the top, while the one corresponding to the relative length of the oyster reef is on the right. To identify the interval bands with better wave attenuation performance, a polynomial fitting line (represented by a dashed line) and its 95% prediction band (represented by the shaded areas of the same color) are included in the scatter plot.

From Cluster 1 to Cluster 6, the fitting lines (the dash-dotted line) gradually shift from the lower right to the upper left corner while increasing their length, i.e. covering a wider range in the parameter space as the 95% prediction band. Moreover, the curvature of the polynomial fitting changes from convex to concave. Observing the top probability distribution, the peak gradually decreases, and its corresponding relative submergence moves downstream. This indicates that when the relative submergence is larger, both Cluster 1 and Cluster 6 are present. In other words, even if a specific range of values (as shown in the enlarged plot) is met, data from other clusters can still be within this optimal interval, leading to larger transmission coefficients, particularly those exceeding 0.5.

We now focus our attention in the optimal parameter space shown as zoom rectangle box in Fig. 12, and we analyze these subspace considering the number of reefs N and the probability associated to each cluster. The results of this analysis is shown in Fig. 13.

Assuming the relative submergence and the length of the oyster reef fall within the optimal interval zone, the range of the transmission coefficient varies significantly with different numbers of oyster reefs. When N is 4, Cluster 1 accounts for 70.45%, far exceeding the probability associated to the other clusters. For $N = 3$, Cluster 1 and Cluster 2 each account for 44.90%. For $N = 2$, the proportions are mainly concentrated in Cluster 2 and Cluster 3. When N is 1, the probabilities of Cluster 4 and Cluster 5 are much higher than the rest of the clusters. As the number of oyster reefs increases, the interval band with the highest probability of occurrence gradually shifts upward (i.e., the cluster number increases). Even within the same optimal interval band, having more oyster reefs makes it easier to achieve lower transmission coefficients. Therefore, although there are different configurations of transmission coefficients within this interval, by considering the number of oyster reefs number, we can estimate the probability of obtaining a lower transmission coefficient.

From the above discussion, we are able to suggest general criteria to optimize the wave attenuation of artificial submerged oyster reefs. The same principle could guide oyster reef restoration activities. It is crucial to set the geometry of the oyster reef in terms of d/h, $kL/(2\pi)$ and N to meet specific conditions. In particular, it is desirable that these parameters fall within the effective interval band as suggested by the clustering analysis and finally reported in Fig. 14 panel (a) as red polygon. Additionally, it is important to consider the number of oyster reefs and its influence on the wave dissipation performance, as summarized in Fig. 14 panel (b). For example, when the number of oyster reefs is set to 4, the probability of the oyster reef transmission coefficient being within the range of $[0.19, 0.51]$ is 70.45%, whereas the probability within a larger range of $K_t \in [0.19, 0.70]$ reaches 89.80%. Generally, once the number of oyster reefs is fixed, a transmission coefficient range with a higher probability of nearly 90% can be determined. These results enable rapid estimation of wave dissipation performance based on the prescribed configuration and provide an optimal strategy for practical oyster reef restoration projects.

It is worth noting that using a larger dataset of simulations (set 2) lead to very similar conclusions in terms of optimal parameters, see Fig. S4 of the Supplementary Information document. One significant difference can be observed in the probability for each cluster for varying reef number, compare Fig. 13 and Fig. S3 of the SI document. The larger number of simulations redistribute the probability linked to the single cluster, however, without changing the main outputs of the analysis.

4.4. Limitations and future directions

Finally, we acknowledge some limitations of the present approach and selected parameters that in the future could be overcome. A first

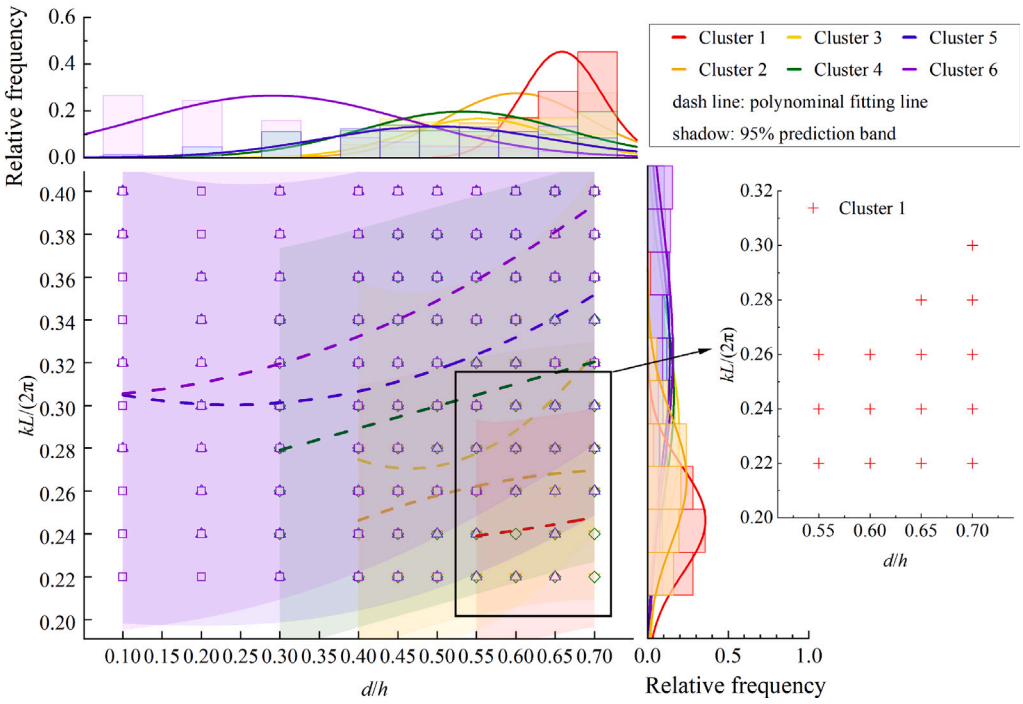

Fig. 12. Bivariate plot and frequency distributions considering the combined effect of length ($kL/(2\pi)$) and submergence (d/h) of oyster reefs in different clusters. Dashed lines indicate the polynomial fitting with the corresponding 95% prediction band as shaded colored areas. The rectangle box corresponding to Cluster 1 is enlarged on the right scatter plot.

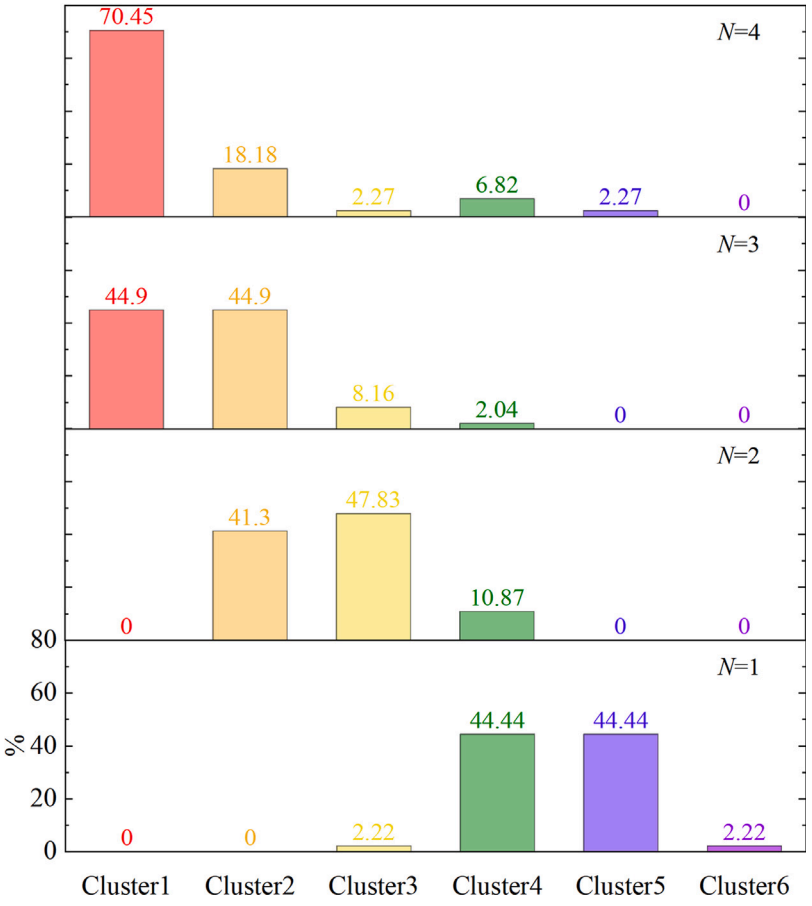

Fig. 13. Influence of number N of oyster reefs on the transmission coefficient, assuming the optimal conditions in terms of d/h.

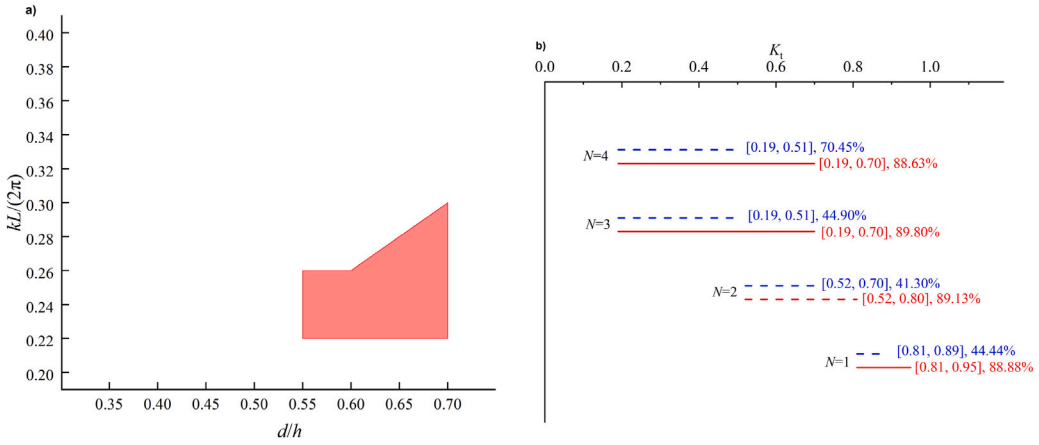

Fig. 14. Optimal strategy of oyster reef restoration. (a) Illustration of effective subdomain in the space $(kL/(2\pi), d/h)$; (b) Transmission coefficient range with different N.

limitation regards the wave model here implemented. If on one hand High-order Spectral (HOS) models offer the possibility to efficiently perform a large volume of simulations with high accuracy; on the other hand, they are limited to non-breaking conditions. To cover a wider range of wave parameters that include braking, a different class of models should be selected. However, using more refined numerical models, e.g. based on Reynolds-averaged Navier–Stokes equations (RANS) equations would limit the number of simulation due to a much higher numerical time costs.

While the HOS method excels in simulating broad-banded wave fields with moderate nonlinearity (Dommermuth and Yue, 1987), it faces challenges in resolving wave breaking and turbulent dissipation due to its spectral truncation of high-order terms (Craig et al., 2021). Recent advances in hybrid modeling (Ducrozet et al., 2016) suggest that combining HOS with conformal mapping or VOF (Volume-of-Fluid) methods could leverage their respective advantages: HOS for open-sea propagation and conformal mapping for near-shore/breaking regimes, for instance, the conformal mapping method (Viotti et al., 2014) provides a truly fully nonlinear formulation in two dimensions by exploiting analytical continuation via the Cauchy-Riemann equations. This approach eliminates truncation errors in vertical dynamics, making it particularly suited for extreme waves and breaking processes, to guide future research in this direction. Current HOS implementation neglects turbulent dissipation mechanisms, which may lead to an overestimation of wave energy retention in breaking scenarios. To address this, future studies could integrate LES (Large Eddy Simulation) closures (Derakhti et al., 2020, among others) or adopt conformal mapping-based Eulerian tracking of vorticity. Based on the above improvements, a promising extension in future work is to hybridize the present HOS framework with conformal mapping techniques, which has the potential to be available for nearshore applications.

A second, important limitation of the present study is the inclusion of the porosity of the submerged reefs. Several studies show that the porosity of the submerged structure could strongly impact the wave transmission (van Gent et al., 2023; Huang et al., 2024; Provan et al., 2024; Huang et al., 2025, among others). This aspect will be considered in the future, implementing porous boundary conditions within the HOS model. A final limitation could be the shape of the reef here considered. Indeed, the shape of the submerged structure is known to play a role in wave transmission.

Regardless these limitations, the main conclusions of the study remains valid in the range of parameters considered that are commonly found in realistic projects. natural and artificial oyster reefs and the related restoration projects as Nature-based Solutions are now considered a valuable strategy to increase coastal climate adaptations. Their use is mentioned on several international guidelines and climate change reports (IUCN, 2020; Wild et al., 2020; Pörtner et al., 2022). The present results could help NGOs and practitioner in the design of these solutions to increase the coastal protection against wave storms impact. Moreover, the suggested approach is easily extended to include more sophisticated wave models or overcoming the limitations in the wave parameters and oyster reefs characteristics (shape and porosity).

Future improvement of the present study will involve laboratory experiments with the aim to further confirm the numerical results and, possibly extend the wave parameters up to breaking conditions. More complex modeling approaches could be used with the aim to include other important processes, such as sediment transport and biological processes typical of living shoreline (e.g. growth of new oyster communities on the artificial reefs).

5. Conclusions

In the present study we conducted an extensive numerical simulation analysis of the effect of submerged reefs on wave transmission, taking advantage of the computational efficiency of the High-order Spectral (HOS) models, validated against experimental data.

We accounted for a large set of controlling parameters (seven variables), the combination of which has been optimized using D-optimal strategy. The use of D-optimal represented an important novelty of the proposes approach leading to an important reduction in terms of number of necessary simulations. The large dataset of results in terms of transmission coefficient has been analyzed implementing a weighted K-means clustering algorithm. The application of the K-means successfully lead to the identification of a relatively small number of clusters, well separated in the physical parameter space and with a small SSE. The results suggested that three dimensionless variables, namely, height of oyster reef relative to water depth (d/h), length of oyster reef relative to wavelength ($kL/(2\pi)$), number of oyster reefs (N), mainly control the wave attenuation process compared to the other variables investigated. The statistical analysis of each cluster allowed for the identification of the optimal combination of the controlling variables. A range of values has finally been suggested to guide the design of new oyster reef or restoration projects.

CRediT authorship contribution statement

Lei Wang: Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Weikai Tan:** Writing – original draft, Software, Data curation, Conceptualization. **Marine Thomas:** Writing – review & editing, Conceptualization. **Felix Leung:** Writing – review & editing, Conceptualization. **Alessandro Stocchino:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

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.

Acknowledgments

The work described in this paper was partially supported by the Research Grants Council of the Hong Kong Special Administrative Region, China (Project Reference Number: AoE/P-601/23-N and project Reference Number: T22-607-24N) and also support by The Nature Conservancy Hong Kong Foundation.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.coastaleng.2025.104751.

Data availability

Data will be made available on request.

References

Abascal, A.J., Castanedo, S., Medina, R., Losada, I.J., Alvarez-Fanjul, E., 2009. Application of HF radar currents to oil spill modelling. Marine Poll. Bull. 58, 238–248.

Ackerman, M., Ben-David, S., Brânzei, S., Loker, D., 2021. Weighted clustering: Towards solving the user's dilemma. Pattern Recognit. 120, 108152.

de Aguiar, P.F., Bourguignon, B., Khots, M., Massart, D., Phan-Than-Luu, R., 1995. D-optimal designs. Chemometr. Intell. Lab. Syst. 30, 199–210.

Ba-Abbad, M.M., Kadhum, A.A.H., Mohamad, A.B., Takriff, M.S., Sopian, K., 2013. Optimization of process parameters using d-optimal design for synthesis of zno nanoparticles via sol–gel technique. J. Ind. Eng. Chem. 19, 99–105.

Bailard, J.A., DeVries, J.W., Kirby, J.T., 1992. Considerations in using bragg reflection for storm erosion protection. J. Waterw., Port, Coast., Ocean. Eng. 118, 62–74.

Baine, M., 2001. Artificial reefs: A review of their design. Appl., Manag. Perform.. Ocean. & Coast. Manag. 44, 241–259.

Beck, M.W., Brumbaugh, R.D., Airoldi, L., Carranza, A., Coen, L.D., Crawford, C., Defeo, O., Edgar, G.J., Hancock, B., Kay, M.C., et al., 2011. Oyster reefs at risk and recommendations for conservation,restoration, and management. Biosci. 61, 107–116.

Beji, S., Battjes, J., 1993. Experimental investigation of wave propagation over a bar. Coast. Eng. 19, 151–162.

Berger, J., Dutykh, D., Mendes, N., 2017. On the optimal experiment design for heat and moisture parameter estimation. Exp. Therm. Fluid Sci. 81, 109–122.

Bodunwa, O., Adewole, A., 2022. A-and d-optimal designs in real life data using imperialist comparative algorithm. Ghana J. Sci. 63, 12–20.

Bowden, G.D., Pichler, B.J., Maurer, A., 2019. A design of experiments (doe) approach accelerates the optimization of copper-mediated 18f-fluorination reactions of arylstannanes. Sci. Rep. 9 (11370).

Bragg, W.H., Bragg, W.L., 1913. The reflection of x-rays by crystals. Proc. R. Soc. Lond.. Ser. A 88, 428–438.

Bredes, A.L., Miller, J.K., Kerr, L., Brown, D.R., 2022. Observations of wave height amplification behind an oyster castle breakwater system in a high-energy environment: Gandys beach NJ. Front. Built Environ. 8, 884795.

Bridges, T., King, J., Simm, J.D., Beck, M., Collins, G., Lodder, Q., Mohan, R., 2021. International Guidelines on Natural and Nature-Based Features for Flood Risk Management. USACE.

Brown, L.A., Furlong, J.N., Brown, K.M., La Peyre, M.K., 2014. Oyster reef restoration in the northern gulf of mexico: Effect of artificial substrate and age on nekton and benthic macroinvertebrate assemblage use. Restoration Ecol. 22, 214–222.

Buccino, M., Calabrese, M., 2007. Conceptual approach for prediction of wave transmission at low-crested breakwaters. J. Waterw. Port Coast. Ocean. Eng. 133, 213–224.

Cawley, K., 2019. When to Use Supervised and Unsupervised Data Mining.

Chang, H.K., Liou, J.C., 2007. Long wave reflection from submerged trapezoidal breakwaters. Ocean Eng. 34, 185–191.

Chong, B., et al., 2021. K-Means Clustering Algorithm: A Brief Review, Vol. 4. pp. 37–40.

Chowdhury, M.S.N., La Peyre, M., Coen, L.D., Morris, R.L., Luckenbach, M.W., Ysebaert, T., Walles, B., Smaal, A.C., 2021. Ecological engineering with oysters enhances coastal resilience efforts. Ecol. Eng. 169, 106320.

Coen, L.D., Brumbaugh, R.D., Bushek, D., Grizzle, R., Luckenbach, M.W., Posey, M.H., Powers, S.P., Tolley, S.G., 2007. Ecosystem services related to oyster restoration. Mar. Ecol. Prog. Ser. 341, 303–307.

Coen, L.D., Luckenbach, M.W., 2000. Developing success criteria and goals for evaluating oyster reef restoration: Ecological function or resource exploitation?. Ecol. Eng. 15, 323–343.

Craig, W., Guyenne, P., Sulem, C., 2021. Normal form transformations and dysthe's equation for the nonlinear modulation of deep-water gravity waves. Water Waves 3, 127–152.

Cremonini, G., De Leo, F., Stocchino, A., Besio, G., 2021. On the selection of time-varying scenarios of wind and ocean waves: Methodologies and applications in the north tyrrhenian sea. Ocean. Model. 163, 101819.

Cuervo, D.Palhazi, Goos, P., Sörensen, K., 2016. Optimal design of large-scale screening experiments: A critical look at the coordinate-exchange algorithm. Stat. Comput. 26, 15–28.

Daliri, M., De Leo, F., Loarca, A.M.L., Scovenna, M., Stocchino, A., Capello, M., Cutroneo, L., Besio, G., 2025. From hindcast to forecast: A statistical framework for real-time coastal circulation bulletins in the gulf of Genoa. Appl. Ocean Res. 154, 104337.

d'Angremond, K., Van Der Meer, J.W., De Jong, R.J., 1996. Wave transmission at low-crested structures. Coast. Eng. 1996, 2418–2427.

Derakhti, M., Kirby, J.T., Banner, M.L., Grilli, S.T., Thomson, J., 2020. A unified breaking onset criterion for surface gravity water waves in arbitrary depth. J. Geophys. Res. Ocean. 125, e2019JC015886.

Dommermuth, D.G., Yue, D.K., 1987. A high-order spectral method for the study of nonlinear gravity waves. J. Fluid Mech. 184, 267–288.

Ducrozet, G., Bonnefoy, F., Le Touzé, D., Ferrant, P., 2012. A modified high-order spectral method for wavemaker modeling in a numerical wave tank. Eur. J. Mech. B Fluids 34, 19–34.

Ducrozet, G., Bonnefoy, F., Le Touzé, D., Ferrant, P., 2016. Hos-ocean: Open-source solver for nonlinear waves in open ocean based on high-order spectral method. Comput. Phys. Comm. 203, 245–254.

Faber, V., et al., 1994. Clustering and the continuous k-means algorithm. Los Alamos Sci. 22 (67).

Gao, J., Hou, L., Liu, Y., Shi, H., 2024. Influences of bragg reflection on harbor resonance triggered by irregular wave groups. Ocean Eng. 305, 117941.

van Gent, M.R., Buis, L., van den Bos, J.P., Wüthrich, D., 2023. Wave transmission at submerged coastal structures and artificial reefs. Coast. Eng. 184, 104344.

Goda, Y., Suzuki, Y., 1976. Estimation of incident and reflected waves in random wave experiments. Coast. Eng. 1976, 828–845.

Goelz, T., Vogt, B., Hartley, T., 2020. Alternative substrates used for oyster reef restoration: A review. J. Shellfish. Res. 39, 1–12.

Grabowski, J.H., Peterson, C.H., 2007. Restoring oyster reefs to recover ecosystem services. Ecosyst. Eng.: Plants Protists 4, 281–298.

Graham, P.M., Palmer, T.A., Pollack, J.Beseres, 2017. Oyster reef restoration: Substrate suitability may depend on specific restoration goals. Restoration Ecol. 25, 459–470.

Guerrero, A., Stephen, K., 2018. Uncertainty analysis using design of experiments to assess the development plan of a north sea field. In: SPE Trinidad and Tobago Section Energy Resources Conference. SPE, D021S011R002.

Haddaji, C., Ennaciri, K., Driouich, A., Digua, K., Souabi, S., 2022. Optimization of the coagulation-flocculation process for vegetable oil refinery wastewater using a full factorial design. Process. Saf. Environ. Prot. 160, 803–816.

Hao, J., Li, J., Liu, S., Wang, L., 2022. Wave amplification caused by bragg resonance on parabolic-type topography. Ocean Eng. 244, 110442.

Hawkins, D., Lye, L.M., 2006. Use of doe methodology for investigating conditions that influence the tension in marine risers for fpso ships. In: First International Structural Specialty Conference. ha06.

Hernández, A.Bersoza, Brumbaugh, R.D., Frederick, P., Grizzle, R., Luckenbach, M.W., Peterson, C.H., Angelini, C., 2018. Storing the eastern oyster: how much progress has been made in 53 years? Front. Ecol. the Environ. 16, 463–471.

Hinton, G.E., Roweis, S., 2002. Stochastic neighbor embedding. Adv. Neural Inf. Process. Syst. 15.

Huan, X., Jagalur, J., Marzouk, Y., 2024. Optimal experimental design: Formulations and computations. Acta Numer. 33, 715–840.

Huang, J., Lowe, R.J., Ghisalberti, M., Hansen, J.E., 2024. Wave transformation across impermeable and porous artificial reefs. Coast. Eng. 189, 104488.

Huang, J., Lowe, R.J., Ghisalberti, M., Hansen, J.E., 2025. Wave dissipation induced by flow interactions with porous artificial reefs. Coast. Eng. 197, 104688.

Huang, J.Z., Ng, M.K., Rong, H., Li, Z., 2005. Automated variable weighting in k-means type clustering. IEEE Trans. Pattern Anal. Mach. Intell. 27, 657–668.

Islam, M.F., Lye, L., 2009. Combined use of dimensional analysis and modern experimental design methodologies in hydrodynamics experiments. Ocean Eng. 36, 237–247.

IUCN, 2020. Guidance for using the iucn global standard for nature-based solutions. a user-friendly framework for the verification. In: Design and Scaling Up of Nature-Based Solutions, first ed. http://dx.doi.org/10.2305/IUCN.CH.2020.09.en.

Iwundu, M., Israel, C., 2024. Performance evaluation of custom a-,d-, and i-optimal designs for non-standard second-order models. Am. J. Theor. Appl. Stat. 13, 92–114. http://dx.doi.org/10.11648/j.ajtas.20241305.11.

Jankovic, A., Chaudhary, G., Goia, F., 2021. Designing the design of experiments (doe)–an investigation on the influence of different factorial designs on the characterization of complex systems. Energy Build. 250, 111298.
Jones, B., Allen-Moyer, K., Goos, P., 2021. A-optimal versus d-optimal design of screening experiments. J. Qual. Technol. 53, 369–382.
Kabagire, K.D., Yahia, A., 2018. Modelling the properties of pervious concrete using a full-factorial design. Road Mater. Pavement Des. 19, 1–17.
Kellogg, M.L., Cornwell, J.C., Owens, M.S., Paynter, K.T., 2013. Denitrification and nutrient assimilation on a restored oyster reef. Mar. Ecol. Prog. Ser. 480, 1–19.
Kim, J.H., Lee, H.C., Kim, J.H., Choi, Y.S., Yoon, J.Y., Yoo, I.S., Choi, W.C., 2015. Improvement of hydrodynamic performance of a multiphase pump using design of experiment techniques. J. Fluids Eng. 137, 081301.
Koszalka, I.M., LaCasce, J.H., 2010. Lagrangian analysis by clustering. Ocean. Dyn. 60, 957–972.
La Peyre, M.K., Humphries, A.T., Casas, S.M., La Peyre, J.F., 2014. Temporal variation in development of ecosystem services from oyster reef restoration. Ecol. Eng. 63, 34–44.
Lagomarsino-Oneto, D., De Leo, A., Stocchino, A., Cucco, A., 2024. Unraveling the non-homogeneous dispersion processes in ocean and coastal circulations using a clustering approach. Geophys. Res. Lett. 51, e2023GL107900.
Larsen, J., Dancy, H., 1983. Open boundaries in short wave simulations—a new approach. Coast. Eng. 7, 285–297.
Lau, S.C., Thomas, M., Hancock, B., Russell, B.D., 2020. Restoration potential of asian oysters on heavily developed coastlines. Restoration Ecol. 28, 1643–1653.
Lee, M.O., Otake, S., Kim, J.K., 2018. Transition of artificial reefs (ars) research and its prospects. Ocean & Coastal Management 154, 55–65.
Lee, J.C., Shin, S.C., Kim, S.Y., 2015. An optimal design of wind turbine and ship structure based on neuro-response surface method. Int. J. Nav. Archit. Ocean. Eng. 7, 750–769.
Li, A.j., Liu, Y., Liu, X., Zhao, Y., 2020. Analytical and experimental studies on bragg scattering of water waves by multiple submerged perforated semi-circular breakwaters. Ocean Eng. 209, 107419.
Lin, P., Liu, H.W., 2005. Analytical study of linear long-wave reflection by a two-dimensional obstacle of general trapezoidal shape. J. Eng. Mech. 131, 822–830.
Liu, H.W., Li, X.F., Lin, P., 2019. Analytical study of Bragg resonance by singly periodic sinusoidal ripples based on the modified mild-slope equation. Coast. Eng. 150, 121–134.
Liu, H.W., Zeng, H.D., Huang, H.D., 2020. Bragg resonant reflection of surface waves from deep water to shallow water by a finite array of trapezoidal bars. Appl. Ocean Res. 94, 101976.
Van der Meer, J.W., Briganti, R., Zanuttigh, B., Wang, B., 2005. Wave transmission and reflection at low-crested structures: Design formulae oblique wave attack and spectral change. Coast. Eng. 52, 915–929.
Mei, C.C., Hara, T., Naciri, M., 1988. Note on bragg scattering of water waves by parallel bars on the seabed. J. Fluid Mech. 186, 147–162.
Melas, V., Nachtmann, G., Rasch, D., 2011. Optimal design of experiments–theory and application. In: Proceedings of the International Conference in Honor of the Late JagdishSrivastava Satellite Conference of the 58th ISI World Statistics Congress. Center of Experimental Design Dublin.
Meyers, L.S., Gamst, G., Guarino, A.J., 2016. Applied Multivariate Research: Design and Interpretation. Sage publications.
Mitsch, W.J., 2012. What is ecological engineering? Ecol. Eng. 45, 5–12.
Modha, D.S., Spangler, W.S., 2003. Feature weighting in k-means clustering. Mach. Learn. 52, 217–237.
Morris, R.L., Heery, E.C., Loke, L.H., Lau, E., Strain, E., Airoldi, L., Alexander, K.A., Bishop, M.J., Coleman, R.A., Cordell, J.R., et al., 2019. Design options, implementation issues and evaluating success of ecologically engineered shorelines. In: Oceanography and Marine Biology.
Morris, R.L., La Peyre, M.K., Webb, B.M., Marshall, D.A., Bilkovic, D.M., Cebrian, J., McClenachan, G., Kibler, K.M., Walters, L.J., Bushek, D., et al., 2021. Large-scale variation in wave attenuation of oyster reef living shorelines and the influence of inundation duration. Ecol. Appl. 31, e02382.
Myers, R.H., Montgomery, D.C., Anderson-Cook, C.M., 2016. Response Surface Methodology: Process and Product Optimization using Designed Experiments. John Wiley & Sons.
Peduzzi, P., Chatenoux, B., Dao, H., De Bono, A., Herold, C., Kossin, J., Mouton, F., Nordbeck, O., 2012. Global trends in tropical cyclone risk. Nat. Clim. Chang. 2, 289–294.
Pierson, K.J., Eggleston, D.B., 2014. Response of estuarine fish to large-scale oyster reef restoration. Trans. Am. Fish. Soc. 143, 273–288.
Pörtner, H.O., Roberts, D.C., Adams, H., Adelekan, I., Adler, C., Adrian, R., Aldunce, P., Ali, E., Begum, R.Ara, Bednar-Friedl, B., et al., 2022. Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Technical Report, Intergovernmental Panel on Climate Change, http://dx.doi.org/10.1017/9781009325844.002.
Provan, M., Rahman, A., Murphy, E., 2024. Full-scale experiments on wave transmission and stability of oyster shell-filled bag berms. Coast. Eng. 193, 104578.
Scyphers, S.B., Powers, S.P., Heck, Jr., K.L., Byron, D., 2011. Oyster reefs as natural breakwaters mitigate shoreline loss and facilitate fisheries. PLoS One 6, e22396.
Seabrook, S.R., Hall, K.R., 1999. Wave transmission at submerged rubblemound breakwaters. Coast. Eng. 1998, 2000–2013.
Seok, W., Kim, G.H., Seo, J., Rhee, S.H., 2019. Application of the design of experiments and computational fluid dynamics to bow design improvement. J. Mar. Sci. Eng. 7 (226).
Singh, M., Xie, W., 2020. Approximation algorithms for d-optimal design. Math. Oper. Res. 45, 1512–1534.
Song, C.Y., Lee, D.J., Lee, J.S., Kim, E.M., Choi, B.Y., 2020. Evaluation of structural design enhancement and sensitivity of automatic ocean salt collector according to design of experiments. J. Ocean Eng. Technol. 34, 253–262.
Srineash, V., Murali, K., 2019. Functional performance of modular porous reef breakwaters. J.Hydro-Enviro. Res. 27, 20–31.
Stokes, S., Wunderink, S., Lowe, M., Gereffi, G., 2012. Restoring gulf oyster reefs. Retrieved from.
Tan, W., Stocchino, A., Cai, Z., 2024. Subspace time series clustering of meteocean data to support ocean and coastal hydrodynamic modeling. Ocean Eng. 313, 119417.
Telford, J.K., 2007. A brief introduction to design of experiments. Johns Hopkins APL Tech. Dig. 27, 224–232.
Thurstan, R.H., McCormick, H., Preston, J., Ashton, E.C., Bennema, F.P., Cetinić, A.B., Brown, J.H., Cameron, T.C., Costa, F.Da, Donnan, D.W., et al., 2024. Records reveal the vast historical extent of european oyster reef ecosystems. Nat. Sustain. 1–11.
Tolley, S.G., Volety, A.K., 2005. The role of oysters in habitat use of oyster reefs by resident fishes and decapod crustaceans. J. Shellfish Res. 24, 1007–1012.
Tsai, L.H., Kuo, Y.S., Lan, Y.J., Hsu, T.W., Chen, W.J., 2011. Investigation of multiply composite artificial bars for bragg scattering of water waves. Coast. Eng. J. 53, 521–548.
Uti, M.N., Din, A.H.M., Yusof, N., Yaakob, O., 2023. A spatial–temporal clustering for low ocean renewable energy resources using k-means clustering. Renew. Energy 219, 119549.
Viotti, C., Dutykh, D., Dias, F., 2014. The conformal-mapping method for surface gravity waves in the presence of variable bathymetry and mean current. Procedia IUTAM 11, 110–118.
Walsh, S.J., Lu, L., Anderson-Cook, C.M., 2024. I-optimal or g-optimal: Do we have to choose? Qual. Eng. 36, 227–248.
Waltham, N.J., Elliott, M., Lee, S.Y., Lovelock, C., Duarte, C.M., Buelow, C., Simenstad, C., Nagelkerken, I., Claassens, L., Wen, C.K., et al., 2020. Un decade on ecosystem restoration 2021–2030—what chance for success in restoring coastal ecosystems? Front. Mar. Sci. 7, 71.
Wang, X., Feng, J., Lin, C., Liu, H., Chen, M., Zhang, Y., 2022. Structural and functional improvements of coastal ecosystem based on artificial oyster reef construction in the Bohai sea, China. Front. Mar. Sci. 9, 829557.
Wang, S.K., Hsu, T.W., Tsai, L.H., Chen, S.H., 2006. An application of Miles' theory to bragg scattering of water waves by doubly composite artificial bars. Ocean Eng. 33, 331–349.
West, B.J., Brueckner, K.A., Janda, R.S., Milder, D.M., Milton, R.L., 1987. A new numerical method for surface hydrodynamics. J. Geophys. Res. Ocean. 92, 11803–11824.
Wiberg, P.L., Taube, S.R., Ferguson, A.E., Kremer, M.R., Reidenbach, M.A., 2019. Wave attenuation by oyster reefs in shallow coastal bays. Estuar. Coasts 42, 331–347.
Wild, T., Freitas, T., Vandewoestijne, S., 2020. Nature-Based Solutions: State of the Art in EU-Funded Projects. Publications Office of the European Union.
Williams, J.M., Lau, S.C., Thomas, M., Tsao, C.S., Russell, B.D., 2024. A long history: Social-ecological systems as drivers of oyster reef loss in the pearl river delta and the broader asian region. In: Oceanography and Marine Biology. CRC Press, pp. 3–191.
Wong, W.K., 1994. Comparing robust properties of a,d, e and g-optimal designs. Comput. Statist. Data Anal. 18, 441–448.
Xie, J.J., 2022. Long wave reflection by an array of submerged trapezoidal breakwaters on a sloping seabed. Ocean Eng. 252, 111138.
Xie, J.J., Liu, H.W., Lin, P., 2011. Analytical solution for long-wave reflection by a rectangular obstacle with two scour trenches. J. Eng. Mech. 137, 919–930.
Xu, W., Tao, A., Zheng, J., 2023. Protect oyster reefs in china's coastal zone. 380, 142.
Zhang, J.S., Jeng, D.S., Liu, P.F., Zhang, C., Zhang, Y., 2012. Response of a porous seabed to water waves over permeable submerged breakwaters with bragg reflection. Ocean Eng. 43, 1–12.
Zhu, L., Chen, Q., Wang, H., Capurso, W., Niemoczynski, L., Hu, K., Snedden, G., 2020. Field observations of wind waves in upper delaware bay with living shorelines. Estuar. Coasts 43, 739–755.
Zu Ermgassen, P.S., Spalding, M.D., Grizzle, R.E., Brumbaugh, R.D., 2013. Quantifying the loss of a marine ecosystem service: Filtration by the eastern oyster in us estuaries. Estuar. Coasts 36, 36–43.