A Novel Input Schematization Method for Coastal Flooding Early Warning Systems Incorporating Climate Change Impacts

Andreas G. Papadimitriou, Anastasios S. Metallinos, Michalis K. Chondros ⬤ and Vasiliki K. Tsoukala *⬤

Laboratory of Harbour Works, School of Civil Engineering, National Technical University of Athens, 5 Heroon Polytechniou Str., 15780 Zografou, Greece; andrewtnt@mail.ntua.gr (A.G.P.); ametallinos@mail.ntua.gr (A.S.M.); michondros@mail.ntua.gr (M.K.C.)
* Correspondence: tsoukala@mail.ntua.gr

Abstract: Coastal flooding poses a significant threat to coastal communities, adversely affecting both safety and economic stability. This threat is exacerbated by factors such as sea level rise, rapid urbanization, and inadequate coastal infrastructure, as noted in recent climate change reports. Early warning systems (EWSs) have proven to be effective tools in coastal planning and management, offering a high cost-to-benefit ratio. Recent advancements have integrated operational numerical models with machine learning techniques to develop near-real-time EWSs, leveraging data obtained from reputable databases that provide reliable hourly sea-state and sea level data. Despite these advancements, a stepwise methodology for selecting representative events, akin to wave input reduction methods used in morphological modeling, remains undeveloped. Moreover, existing methodologies often overlook the significance of compound extreme events and their potential increased occurrence under climate change projections. This research addresses these gaps by introducing a novel input schematization method that combines efficient hydrodynamic modeling with clustering algorithms. The proposed methodology, implemented in the coastal area of Pyrgos, Greece, aims to select an optimal number of representative sea-state and water level combinations to develop accurate EWSs for coastal flooding risk prediction. A key innovation of this methodology is the incorporation of weights in the clustering algorithm to ensure adequate representation of extreme compound events, also taking into account projections for future climate scenarios. This approach aims to enhance the accuracy and reliability of coastal flooding EWSs, ultimately improving the resilience of coastal communities against imminent flooding threats.

Keywords: coastal flooding; climate change; clustering; numerical modeling; early warning system; input schematization; compound events

1. Introduction

Coastal flooding is an imminent threat for coastal areas, compromising the safety of communities and having adverse impacts on the economy [1] and causing damage to infrastructure. This threat is further intensified due to sea level rise, rapid urbanization of coastal areas, and inadequate protection of coastal structures, as highlighted in the published report of the Intergovernmental Panel on Climate Change [2]. In this context, compound extreme weather events refers to the simultaneous occurrence of extreme wave heights and increased water level elevations due to storm surges and astronomical tides [3–5]. Consequently, the increase in their frequency of occurrence may substantially increase the associated risk [6]. Understanding the mechanisms that drive coastal flooding [7–9] and enhancing the resilience of coastal communities against climate change impacts [10,11] have been a focal point for coastal engineering research in recent years.

To protect livelihoods and properties of coastal communities at risk, administrative efforts should focus on anticipating and alerting citizens in advance to the imminent threat of coastal flooding. To this end, early warning systems (EWSs) are a promising tool for

coastal planning and management, with a demonstrated high cost-to-benefit ratio [12]. Various research efforts have focused on the development of EWSs for coastal flooding risk, utilizing operational numerical modeling predictions [13–15], often combining them with machine learning techniques [16–18]. Machine learning techniques are especially alluring for the development of near-real-time EWSs considering the high computational resources required by hydrodynamic models that are tasked with predicting coastal flooding inundation due to compound extreme events.

Lately, with the prevalence of metocean databases (e.g., [19–21]) which offer reliable predictions of sea-state offshore wave characteristics and sea level data at hourly intervals, there is an abundance of data to be considered in the framework of carrying out hydrodynamic simulations, and subsequently, training a machine learning algorithm. In other coastal engineering applications, the concept of schematizing the forcing input for numerical models is widespread, with the development of various wave input reduction or wave schematization methods [22–26] to accelerate the interannual predictions of coastal bed evolution.

In a similar manner, research efforts that focus on the development of EWSs for coastal flood events usually follow a procedure of discretizing sea-state wave characteristics and water level variations in bins of fixed intervals to reduce the required computational effort. Chondros et al. [16] trained an artificial neural network (ANN) to predict the coastal flooding risk for the coastal area of Rethymno Port in Crete. To train the ANN, the authors utilized a parabolic mild slope wave model [27] for wave propagation to provide boundary conditions for coastal inundation modeling using the HEC-RAS 2D [28] hydrodynamic model. By dividing the significant wave height, peak wave period, and mean wave direction into constant bins of 1 m, 1 s, and 30° intervals, respectively, and considering three water level elevations with a constant incremental step of 0.5 m (lowest, mean, and highest water level), a total of 303 sea states were utilized to train an ANN. Garzon et al. [17] trained a Bayesian network with numerical modeling simulations nesting SWAN [29] and XBeach model [30] to develop an EWS for the coastal area of Praia de Faro, Portugal. Since Bayesian networks cannot extrapolate beyond the parameter space in which they are constructed, the authors schematized the offshore sea-state wave characteristics by dividing the wave heights into bins with 0.5 m intervals and 1 s for the peak wave period. They separately considered synthetic storm events as well as large swell events when training the Bayesian network. Epsejo et al. [14] developed an EWS for coastal flooding of low-lying atoll islands in Tarawa, Kiribati. Differently from the previous studies, the authors utilized the maximum dissimilarity selection algorithm (MDA) [31], which focuses on maximizing the dissimilarity or diversity among the chosen parameters, to define 500 representative sea-state scenarios, also considering the wind speed and wind direction as input variables in their analysis. Thereafter, they performed simulations utilizing the SWAN wave model [29] coupled with ADCIRC [32] to define the threshold for coastal flooding risk to be used in the operational EWS.

In light of the above, to the best of the authors' knowledge a stepwise methodology to select schematized events (i.e., a predefined number of representative events) for EWSs for coastal flooding, similar to wave input reduction methods for morphological modeling applications, has not been realized yet. Moreover, the respective methodologies to select schematized events do not consider the presence of compound extreme events of wave height and water level elevation in the analysis, and exclude climate change projections altogether, which may lead to an insufficient training of the machine learning algorithm and compromise the predictions of the EWS. The scope of this research is to present a novel input schematization method combining efficient hydrodynamic modeling and clustering algorithms to select an optimal number of representative sea-state wave characteristics and water level combinations that can be utilized to develop an accurate EWS for coastal flooding risk prediction. One of the central innovative aspects of the proposed methodology is the attention on compound events, that can potentially lead to the inundation of coastal areas, by introducing weights in the clustering algorithm to ensure a good representation of

extreme compound events in the schematized dataset, as well as including climate change projections for wave characteristics and water levels. The methodology was conceptualized with the intent to be easily replicable in other coastal areas but also to provide a framework for selecting representative events in other physics-based applications (e.g., urban or fluvial flooding EWSs, or EWSs for forest fires).

2. Materials and Methods

2.1. Study Area and Available Data

The area of interest is the coastal zone of Pyrgos, the capital city of the regional unit of Elis in Western Peloponnese, Greece. Pyrgos is the administrative and commercial center of the region, with a population of around 25,000 residents, and is located around 30 km from the archaeological site of Olympia, a prominent tourist attraction. The area historically has been subjected to various flood events both fluvial (due to the presence of the Alfios river) as well as coastal, with notable flood events occurring in February 2012, October 2019, and November 2021, with the latter two exclusively attributed to coastal flooding. Consequently, flood protection in the form of interventions along the riverbed of Alfios, as well as EWSs for prevention and evacuation, have been at the forefront of administrative policies and research endeavors.

The latter lies within the scope of the present research, with the ultimate goal to develop an EWS for coastal flood events. The study area where coastal flooding risk will be assessed extends around 4.5 km from both sides of the Alfios river's mouth. A general outline of the study area is illustrated with a purple rectangle in Figure 1. Interestingly, the hinterland is equipped with a means of physical protection against coastal flooding with sand dunes situated at an average distance of 32 m from the coastline with an average height of 2.3 m above the mean sea level. Two indicative coastal profiles, showing the above-mentioned sand dunes are also shown in Figure 2, utilizing information from a recently conducted topo-bathymetric survey and the DEM file with horizontal spacing of 2 m from the National Cadastre of Greece [33]. It is important to note that if waves overtop the sand dunes, it poses a significant threat due to the proximity of residential areas near the shoreline north of the Alfios river's mouth. Additionally, the dried-out Lake Agoulinitsa to the south is a potential flooding hotspot.

The first step in developing an EWS for coastal flooding is the collection of input data, particularly those related to sea-state wave characteristics and sea level elevation. In this research, data from the historical period, along with a future period incorporating the impacts of climate change, were obtained from the Copernicus Climate Change Service [34]. The future projections of the variables of interest are based on two Representative Concentration Pathway (RCP) scenarios. The first, RCP4.5, corresponds to an optimistic emission scenario, where emissions start declining beyond 2040. The second, RCP8.5, represents a pessimistic scenario, where emissions continue to rise throughout the century. The offshore sea-state wave characteristics (wave height, H_{mo}; peak wave period, T_p; and mean wave direction, MWD) were extracted from the "Ocean surface wave time series for the European coast from 1976 to 2100 derived from climate projections" package, whereas total water level data, containing the contributions of residual storm surges and astronomical tides were extracted from the package "Water level change time series for the European coast from 1977 to 2100 derived from climate projections". In this package, sea level data are obtained through an implementation of the Deltares Global Tide and Surge Model (GTSM) version 3.0 forced with regional climate forcing and sea level rise initial conditions based on the high-resolution regional climate change ensemble developed for Europe within the World Climate Research Program Coordinated Regional Downscaling Experiment (EUROCORDEX) initiative. Wave characteristics are calculated with the pan-European wave initiative, a refined version of the WAM model [35] utilized in the premises of ECMWF. Finally, metocean data were obtained for three distinct periods:

- Time period 1977–2005, hereafter denoted as the historical dataset.

- Time period 2041–2070 with forcing input based on RCP 8.5, hereafter denoted as RCP 8.5 dataset.
- Time period 2071–2100 with forcing input based on RCP 4.5, hereafter denoted as RCP 4.5 dataset.

Figure 1. Overview of the study area, highlighting Alfios river and the outline of the dried-out Agoulinitsa Lake along with two characteristic coastal profile locations.
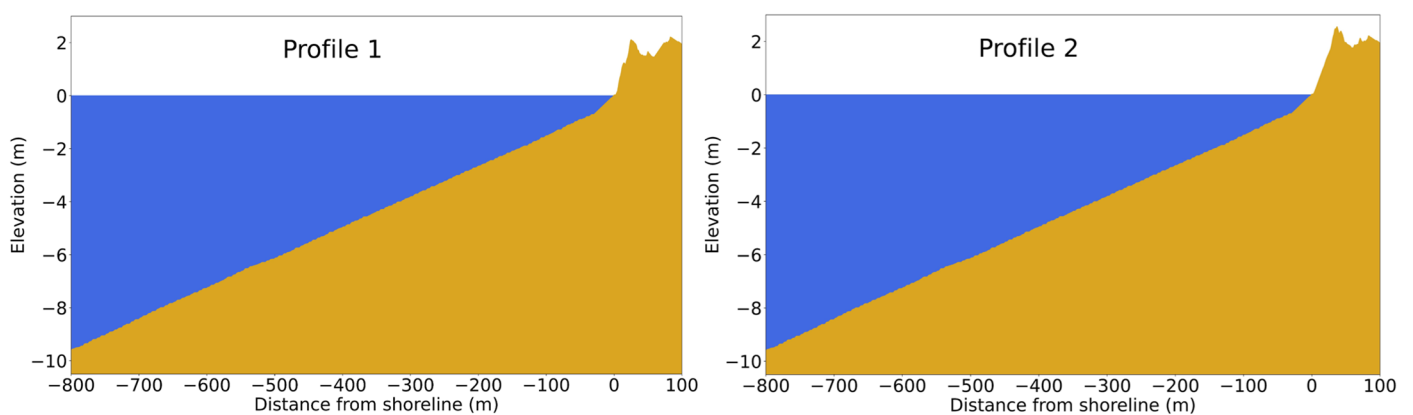
Figure 2. Coastal profile sections for the two examined locations.

Figure 3 presents wave rose plots for each dataset offshore from the study area.
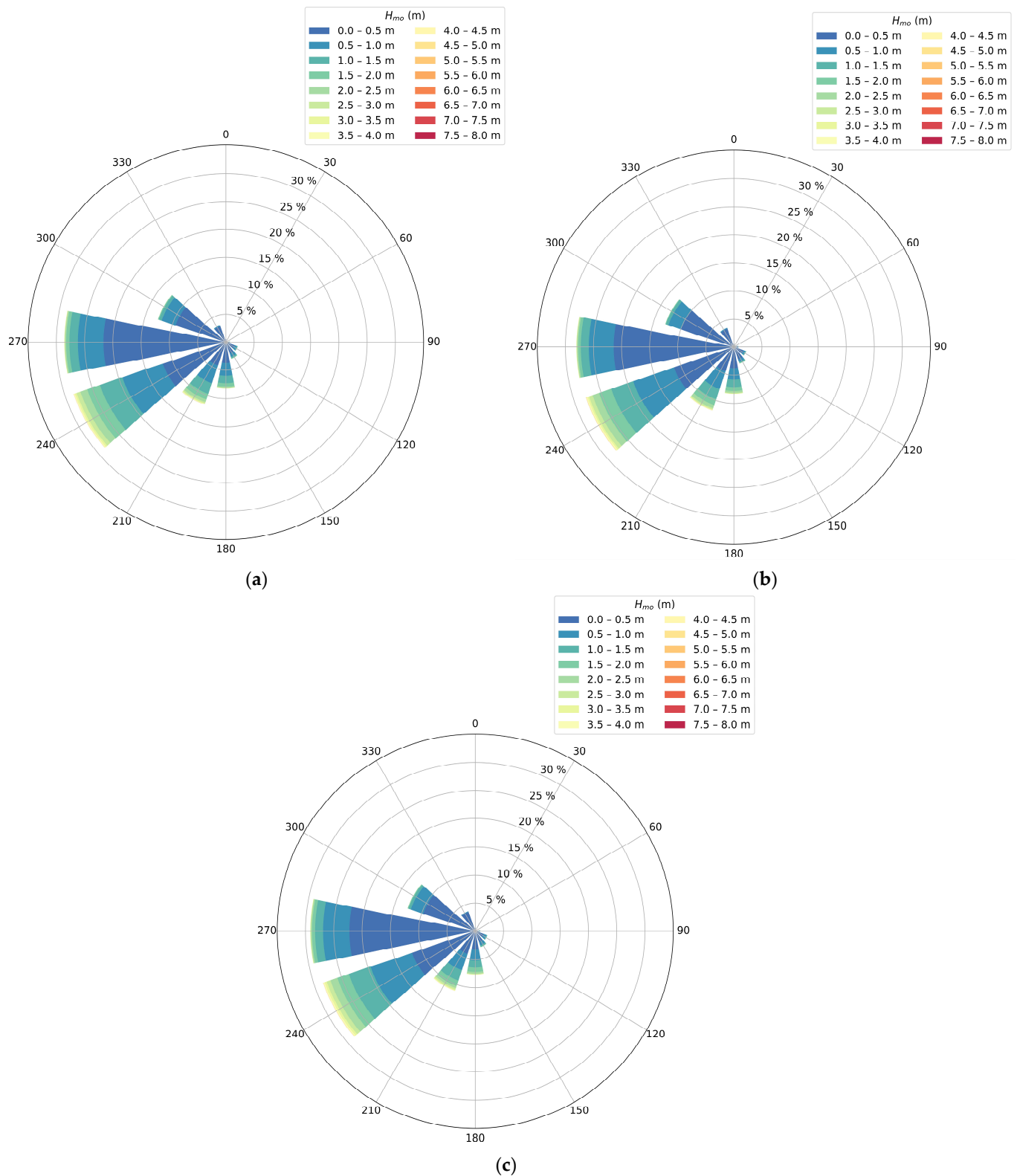
Figure 3. Wave rose plot offshore of the study area for the (**a**) historical, (**b**) RCP 8.5, and (**c**) RCP 4.5 datasets.

As can be seen from Figure 3, the wave regime is similar between the historical period and the two periods corresponding to the RCP 8.5 and RCP 4.5 projections. In particular, two dominant wave directions are identified, namely, from the west (270°) and west-southwest (240°) sectors. Compared to the historical period, the maximum projected

spectral wave height offshore decreases for both RCP scenarios, with values of 7.89 m, 6.24 m, and 6.35 m for the historical, RCP 8.5, and RCP4.5 datasets, respectively. It is noted that waves propagating from the west-southwest direction are not subjected to intense refraction as they propagate nearshore, due to the mean shoreline orientation, since the shore normal is positioned at 225° compared to the true north position.

The total water level elevation, including the contribution of residual storm surges and astronomical tides, along with mean sea level rise for the climate change projections is plotted in Figure 4.
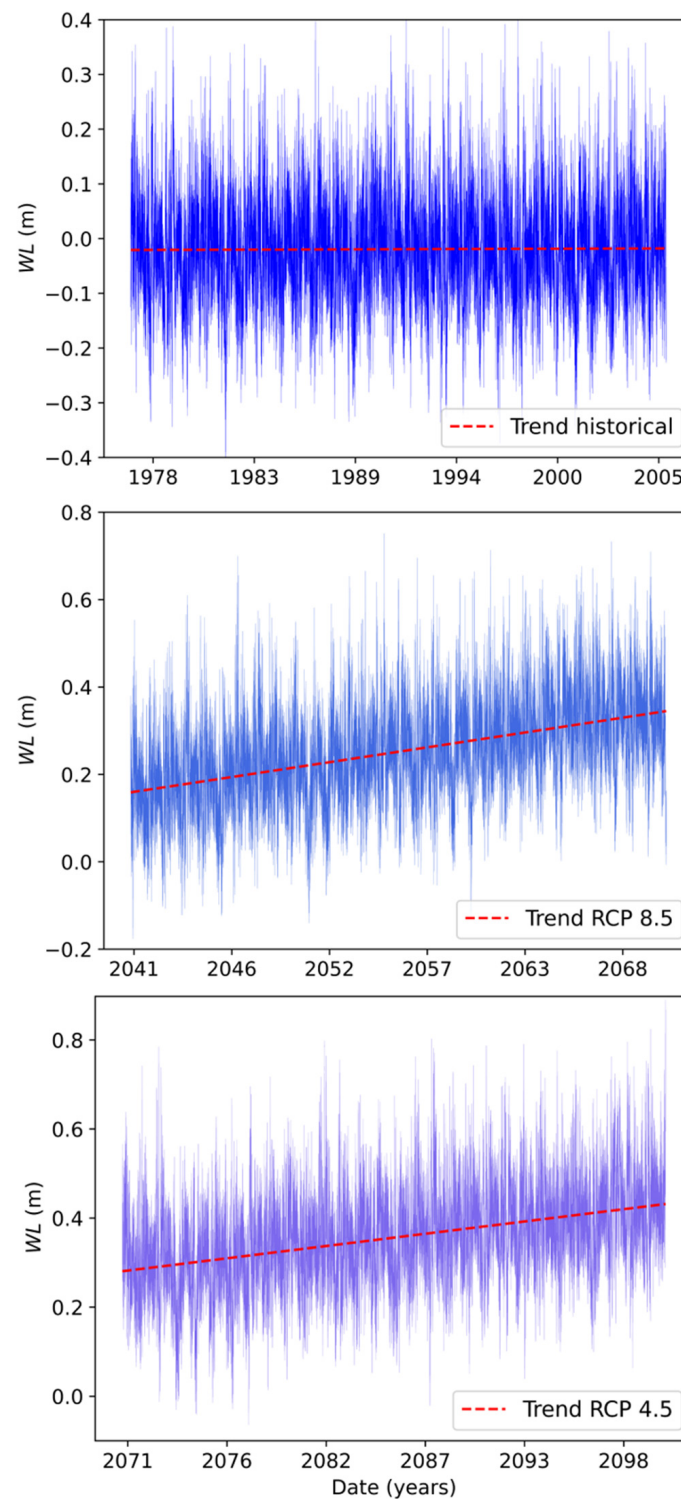
Figure 4. Total water levels for the study area for the historical, RCP 8.5, and RCP 4.5 datasets.

For the historical dataset, the mean water level is located at -0.019 m, with the maximum water level elevation being 0.45 m. As shown in Figure 4, for the period between 1977 and 2005, there is a relatively constant trend for the total water level in the study area. In contrast, for the RCP 8.5 dataset, at the start of 2041 the mean sea level lies at 0.16 m, signifying a sea level rise of about 6 mm/year between the end of the historical period and the start of 2041. From that point, an upward trend is identified for the total water, reaching a maximum projected value of 0.75 m at the end of 2070. Finally, for the RCP 4.5 dataset, covering the period 2071–2100, the mean sea level is located at 0.28 m, signifying a sea level rise of about 4.5 mm/year, with an upward trend for the upcoming years, reaching a value of 0.43 m at the end of the century. The maximum projected value of the total water level elevation for the RCP 4.5 dataset is 0.89 m.

2.2. Methodology Outline

This section presents the outline of the conceptualized methodology to select combinations of wave characteristics based on nearshore wave propagation simulations and machine learning techniques, focusing especially on incorporating the impact of compound flood events. The distinct steps to reproduce the methodology (also illustrated in Figure 5) are the following (and they should be repeated for each period of interest, e.g., historical, future period incorporating climate change):

1. Obtain time series of offshore sea-state wave characteristics (at a minimum H_{mo}, T_p, and MWD should be included) and total water level elevation (WL) for the study area for the historical and future periods incorporating climate change impacts. These data can be retrieved from metocean databases such as the CDS [34].

2. Define compound events of waves and total water level elevations, adopting an extreme value threshold for H_{mo} and WL. This threshold should be reasonably high to capture extreme events and a value of the 95th [36,37] or 99th percentile [38,39]. By adopting the threshold for both variables, an area of compound extremes is defined by both variables exhibiting extreme values at the same time. Hence, two areas are defined: a "supercritical" one, containing the compound events; and a "subcritical" one, lying below the thresholds. It should be noted that if the timestep of the wave characteristic and water level time series is not equal, resampling through extrapolation could be conducted by taking either the mean or maximum value of the variables within the time frame examined.

3. Propagate the defined compound events at a nearshore depth of about 8–10 m (taken to be 10 m in the present study). This depth represents a typical value for the estimation of wave characteristics in the nearshore and in the context of the subsequent calculation of the depth of closure, as shown in other studies [23,40]. For this purpose, it is advised to use either a spectral wave model (e.g., [29,41,42]) or a mild slope wave model [43,44]. In the present study, a nonlinear parabolic mild slope model (hereafter denoted PMS) [26,27] was utilized. The reason for selecting this model is the accuracy in prescribing the wave field in mildly sloping beds, the fast simulation times, and the capability to include the increase in the water level in the wave propagation simulations. It is noted that compound events that have wave incidence angles offshore with $a_o > \pm 90°$, with respect to the shore normal, can be excluded from the analysis as they are not expected to affect the study area and lead to coastal inundation.

4. Calculate wave overtopping across indicative coastal profile sections utilizing the nearshore wave characteristics and water level elevations obtained through step 3. The selected coastal profiles should be representative of the bathymetry and topography of the hinterland. In case of a relatively alongshore homogeneous domain, a single coastal profile can be chosen. Wave overtopping can be calculated either through empirical relationships, developed artificial neural networks [45], or numerical modeling. In the present study, the coastal profile model CSHORE [46] was selected due to its rapid simulation times and capability to predict wave runup and overtopping

in porous media. Model calibration and validation should take place at this stage provided the availability of field measurements.

5. Based on the results of step 4, define events that lead to wave overtopping (with a non-zero wave overtopping discharge or with an overtopping discharge exceeding the tolerable limits, i.e., as presented in [47]). For the set of compound events, perform feature weight estimation [48] to identify the contribution of each variable (i.e., H_{mo}, T_p, MWD, WL) in wave overtopping. In case of negative weights, the corresponding variable can be excluded from the subsequent analysis.

6. Based on the calculated weights, generate a weighting function to quantify the combined contribution of each variable in wave overtopping. In the present research a weighted sum model [49] was adopted, commonly used in multicriteria analysis.

7. Perform weighted cluster analysis separately for the "subcritical" and "supercritical" areas of the dataset. Clustering methods have experienced increased usage in the field of coastal engineering in the past decade to uncover hidden patterns and formations in multivariate datasets. They have been employed in various applications such as coastal storm classification [50], longshore transport rates [51], cross-shore profile morphological evolution [52], multivariate wave climate classification [31], and as input reduction methods for coastal morphological simulations [23]. In the present study, the well-known K-Means algorithm [53] is utilized but other algorithms can be also implemented (e.g., Fuzzy C-Means [54], CURE [55] or the Maximum Dissimilarity Algorithm [31], among others). The reason for the selection of K-Means in the context of the methodology used herein mostly lies in its capability to handle excessively large datasets with computational efficiency. The clustering algorithm takes as input variables H_{mo}, T_p, $\cos(a_o)$, and WL. It is noted that the MWD is transformed to the incident wave angle in relation to the shoreline orientation. Through an iterative procedure the algorithm selects a number of clusters and centroids, the latter of which are the desired representative events. As proposed and supported by the findings of this research, the clustering algorithm should incorporate the weighting function defined in the previous step.

The obtained centroids are the individual events that can be utilized to perform hydrodynamic simulations of coastal inundation, and subsequently, train a machine learning algorithm (e.g., artificial neural networks [16] or Bayesian networks [17]). With the utilization of weighted clustering algorithms, it is foreseen that the subsequent EWS will be properly trained and capable of capturing the possible occurrence of extreme compound flood events, and it will ensure its longevity.

In the following subsection, the governing equations of the two numerical models utilized in the context of the proposed methodology are presented.

2.3. Background of Numerical Models

2.3.1. Nearshore Wave Model

For the simulation of the irregular wave transformation from offshore to nearshore a nonlinear mild-slope wave model of parabolic approximation is implemented, developed in [27], based on the work in [56], who derived a parabolic equation governing the complex amplitude, A, of the fundamental frequency component of a Stokes wave. Thereafter, Ref. [57] improved the parabolic equation and its applicability range through approximations based on minimax principles to allow for large-angle propagation. The range of allowable wave angles within the limitations of the parabolic approximation was significantly increased by relaxing the local accuracy of approximations based on Padé approximants at normal wave incidence in favor of minimax approximations, which minimize the maximum error that occurs over a prespecified range of wave directions. The

revised governing parabolic equation, allowing for the study of waves with larger wave incidence angles with respect to the x axis (the principal direction of propagation), is

$$C_g A_x + i\left(\bar{k} - a_0 k\right) C_g A + \tfrac{1}{2}\left(C_g\right)_x A + \tfrac{i}{\omega}\left(\alpha_1 - b_1 \tfrac{\bar{k}}{k}\right)\left(CC_g A_y\right)_y - \tfrac{b_1}{\omega k}\left(CC_g A_y\right)_{yx}$$
$$+ \tfrac{b_1}{\omega}\left(\tfrac{k_x}{k^2} + \tfrac{\left(C_g\right)_x}{2kC_g}\right)\left(CC_g A_y\right)_y + \tfrac{i\omega k^2}{2} D|A|^2 A + \tfrac{w}{2} A = 0 \tag{1}$$

where the parameter D is given by $D = \dfrac{\left(\cosh 4kh + 8 - 2\tanh^2 kh\right)}{8\sinh^4 kh}$, the complex amplitude A is related to the water surface displacement by $\eta = Ae^{-i(kx - \omega t)}$, and k is the local wave number, related to the angular frequency of the waves, ω, and the water depth, h. Moreover, $\bar{k}$ is a reference wave number taken as the average wave number along the y-axis, C is the phase celerity, C_g is the group celerity, and w is a dissipation factor. Coefficients a_0, α_1, and b_1 depend on the aperture width chosen to specify the minimax approximation [58].
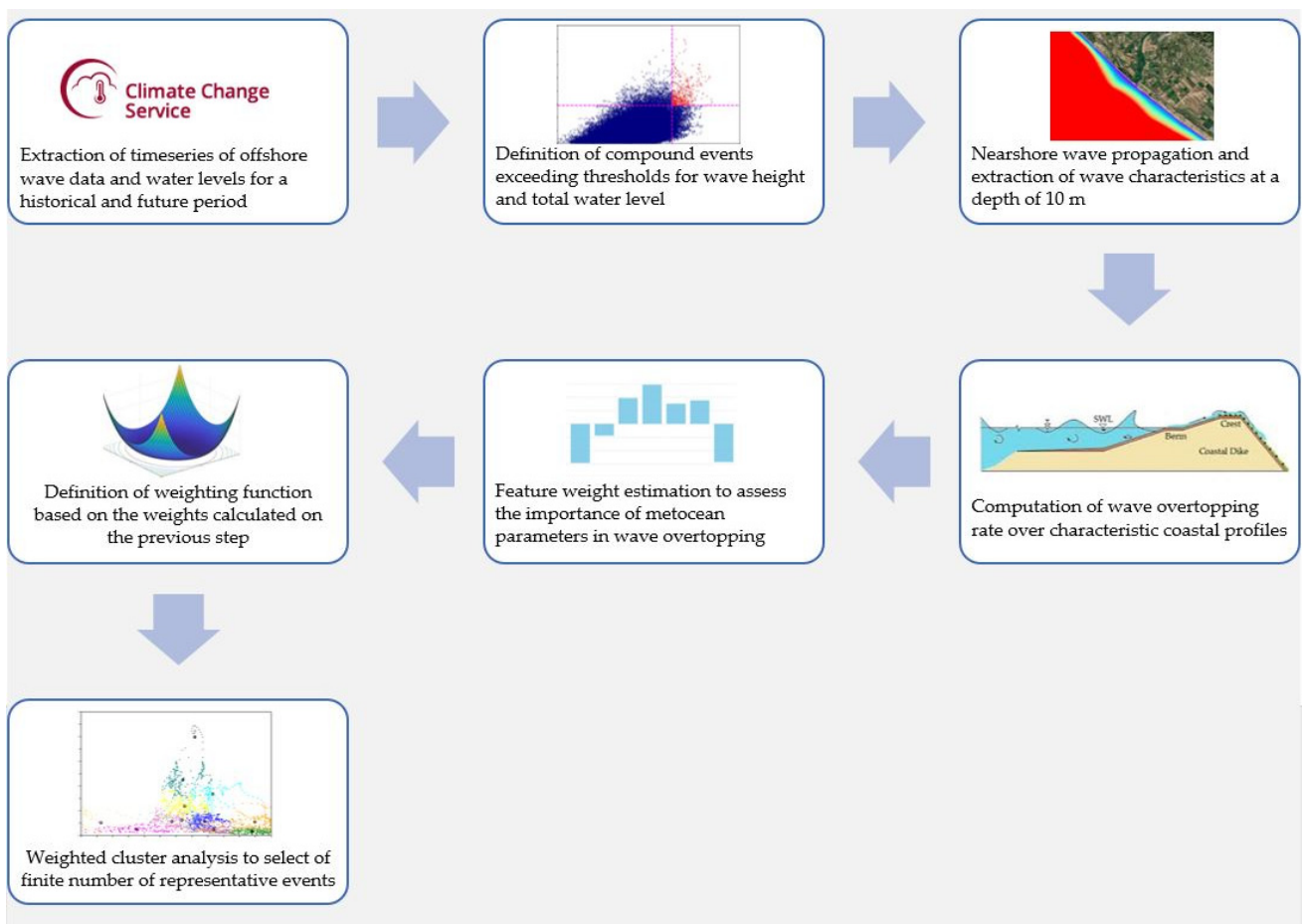

Figure 5. Flow chart depicting the steps of the proposed methodology.

The model allows for the generation and propagation of unidirectional irregular waves by dividing the wave energy spectrum into discrete wave components of equal wave energy and performing separate simulations for each one. Energy dissipation due to bathymetric breaking following the formulation of [59], and bottom friction, which is modeled through the formulation of [60], are also treated by the model. To improve model results in the nearshore, nonlinear dispersion characteristics are incorporated by introducing an approximate nonlinear amplitude dispersion relationship, as shown in [61].

2.3.2. Coastal Profile Model CSHORE

For the simulation of wave overtopping in the nearshore, the coastal profile model CSHORE [46] was utilized. The model is time-averaged and is able to predict the coastal profile evolution at the scale of storm events or interannual periods, as well as the wave runup and overflow due to wave overtopping for permeable structures and beds [62,63] as well as impermeable structures [64,65].

The wave transformation and evolution of the root mean squared wave height (H_{rms}) is obtained by solving the wave action balance equation:

$$\frac{d}{dx}\left[\frac{E}{\omega}\left(C_g\cos\alpha + \frac{Q_x}{\overline{h}}\right)\right] = -\frac{D_b + D_f}{\omega} \tag{2}$$

where Q_x is the cross-shore volume flux, E is the wave energy, a is the wave incidence angle with respect to the cross-shore direction, h is the still water depth, and D_b and D_f are energy dissipation rates due to wave breaking and bottom friction, respectively. In particular, energy dissipation due to wave breaking is calculated based on the formula of [66], modified by [67] as follows:

$$D_b = \frac{\rho g a_s Q_b H_b}{4T_p} \tag{3}$$

where ρ is the seawater density; g is the acceleration due to gravity; T_p is the wave period at the spectral peak; Q_b is the fraction of breaking waves; and $a_s = \frac{2\pi S_b}{3k\overline{h}}$ is a slope effect parameter expressed as the ratio between the wave length ($2\pi/k$) and the horizontal length ($3\overline{h}/S_b$) imposed by the small depth and relatively steep slope.

Energy dissipation due to bottom friction is calculated through the following relationship:

$$D_f = \frac{1}{2}\rho f_b \overline{U}^3 \tag{4}$$

where f_b is a bottom friction coefficient and U is the cross-shore flow velocity.

The cross-shore volume flux Q_x for the flow over an impermeable layer in shallow water is calculated through the time-averaged and vertically-integrated continuity equation as follows [46]:

$$Q_x = \overline{hU} + \frac{g\sigma_\eta^2}{C}\cos\alpha + q_r\cos\alpha = q_o \tag{5}$$

where σ_η is the standard deviation of the free surface elevation, q_r is the roller volume flux, estimated using the roller energy equation as presented in [67,68], and q_o is the wave overtopping rate.

The overtopping rate q_o is estimated using the computed hydrodynamic variables at the crest (highest elevation) of the coastal profile z_b. If the landward marching computation starting from $x = 0$ (i.e., the seaward limit of the profile) does not reach the crest, then the overtopping rate $q_o = 0$.

3. Results and Discussion

3.1. Compound Flood Events

Initially, the basis on which an event is categorized as a compound flood event is laid out. In the context of this study, a compound event is defined as one where the values of the offshore wave height and the total water level exceed the respective 99th percentiles of the corresponding dataset [38,39,69]. Of particular interest is the utilization of three separate datasets (historical, RCP 4.5, and RCP 8.5) covering different time periods. As can be observed from Figure 4, a mean sea level rise due to climate change is identified for both the RCP 8.5 and RCP 4.5 datasets compared to the historical one. It should be noted that the 99th percentiles of total water level for the datasets incorporating climate change are expected to be higher than the respective one for the historical period due to mean sea level

rise. This can lead to the omission of sea level elevations that, combined with high-energy sea states, can potentially lead to overtopping and coastal flooding. Consequently, the 99th percentile values of the total water level and significant wave height were used as constant thresholds for both the historical and the RCP datasets. Figure 6 illustrates the compound events for the three time periods of interest.

It is important to note that the corresponding thresholds define two distinct areas, a "supercritical" area, corresponding to the compound extreme events (illustrated with orange markers in Figure 6), and a "subcritical" area, lying below the thresholds (illustrated with blue markers in Figure 6). These two areas effectively create two separate datasets, which are treated differently in the subsequent cluster analysis.

In total, 475 compound events were obtained for the historical period, whereas the corresponding number was 2572 for RCP 8.5 and 3018 for RCP 4.5, signifying a significant increase in the frequency of future extreme events offshore of the study area, when adopting the thresholds of the historical dataset. A slight change in the mean deviation of the wave incidence angle with respect to the shore normal is observed, with the absolute value being $15.84°$ for the historical dataset, $18.89°$ for RCP4.5, and $17.52°$ for RCP8.5. This indicates that there is slight altering of the mean wave direction, and for the datasets incorporating climate change, more oblique incidence angles are observed, which can potentially be beneficial in reducing wave runup and overtopping of the sand dunes. Even though this shift in wave incidence angles enhances coastal resilience by reducing the extreme wave impacts on the shore, the simultaneous increased water levels under climate change projections could potentially compromise it.

3.2. Coastal Wave Propagation

The next step concerns the coastal wave propagation from offshore to nearshore in order to obtain wave characteristics across the depth of 10 m. For this purpose, a 2D nonlinear irregular parabolic mild slope (PMS) wave propagation model was utilized to conduct the required simulations of wave propagation. The PMS wave model is mainly tasked with producing output that is used to force the CSHORE profile mode. It is advised to utilize a 2D area numerical model for this task, to better prescribe the wave transformation processes before proceeding with the 1D coastal profile modeling.

The bathymetry of the domain was discretized with a spatial step size of 5 m in both directions. The domain extends 13 km in the x-direction and 8.5 km in the y-direction. The points where characteristics were extracted have a depth of 10 m. The bathymetric grid, superimposed with the positions where wave characteristics were extracted is shown in Figure 7. Points P1 and P2 are located at distances of 315 m and 300 m from the shoreline, respectively. This distance is over three times larger than the average wavelength of the incident waves at this depth (assuming linear dispersion of water waves), which is 99.2 m, allowing for a good resolution of the wave transformation processes from this depth up to the shoreline. It is also noted that these two positions were selected for the particular study area due to the presence of the sand dunes, the potential breach of which can lead to coastal flooding, in conjunction with the conducted surveys in the hinterland, which indicate a relatively uniform coastal profile for about 1.5 km on either side of the river mouth. The steps of the methodology can however readily be applied to any number of nearshore locations and coastal profiles, provided significant variations in the coastal profile are identified.
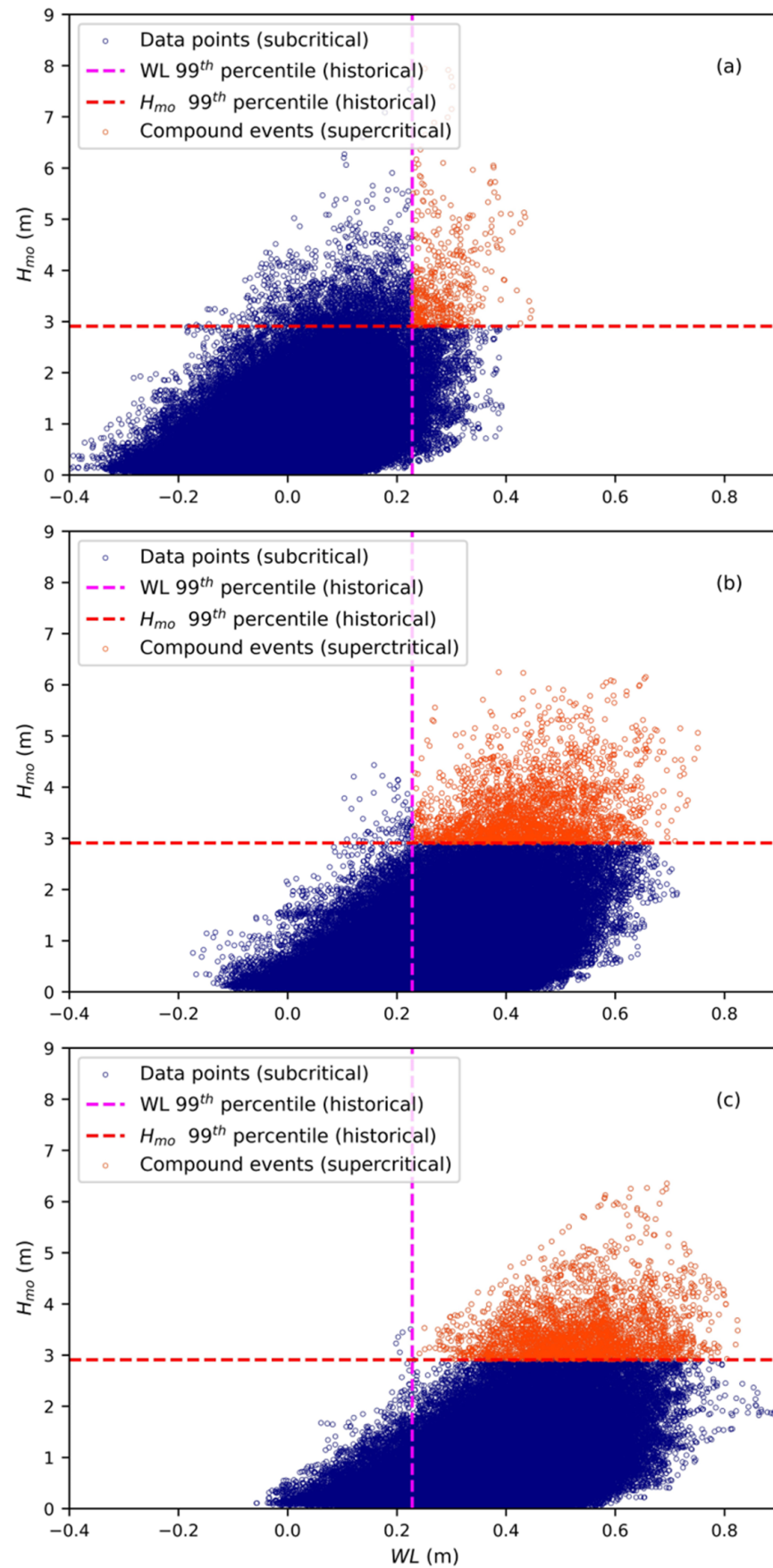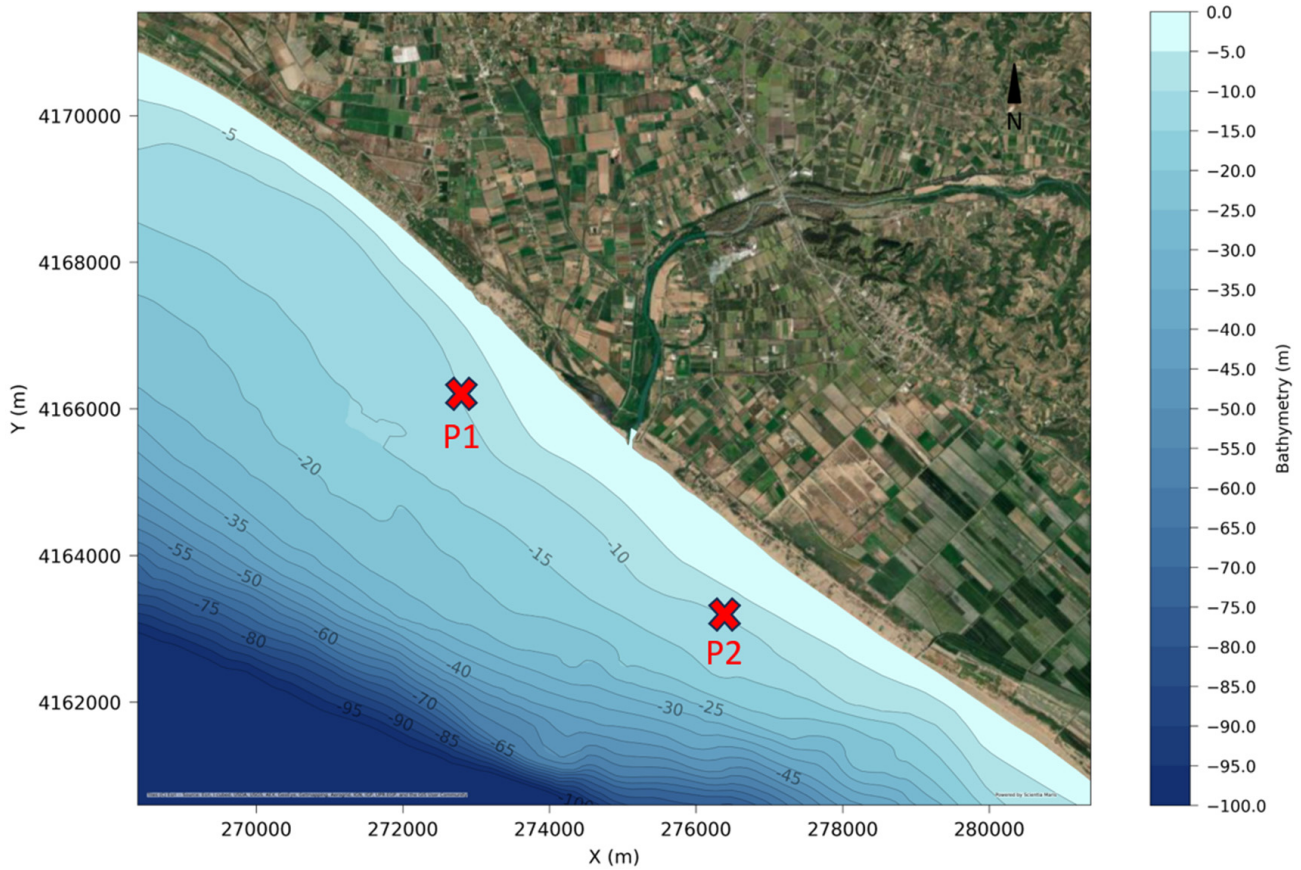
Figure 6. Compound events as defined for the (**a**) historical, (**b**) RCP 8.5, and (**c**) RCP 4.5 datasets.

Figure 7. Bathymetry of the numerical domain and extraction points of nearshore waves. Date of bathymetric survey: 26 July 2024.

The maximum depth is 100 m and the mean coastal bed slope close to the mouth of the Alfios river is relatively mild, reaching a value of 0.8%. Figure 8 shows a comparison between the offshore sea-state wave characteristics and the nearshore values at the two extraction locations (P1 and P2) for the historical dataset, while Figures 9 and 10 correspond to the RCP8.5 and RCP4.5 datasets, respectively, as simulated with the PMS model.
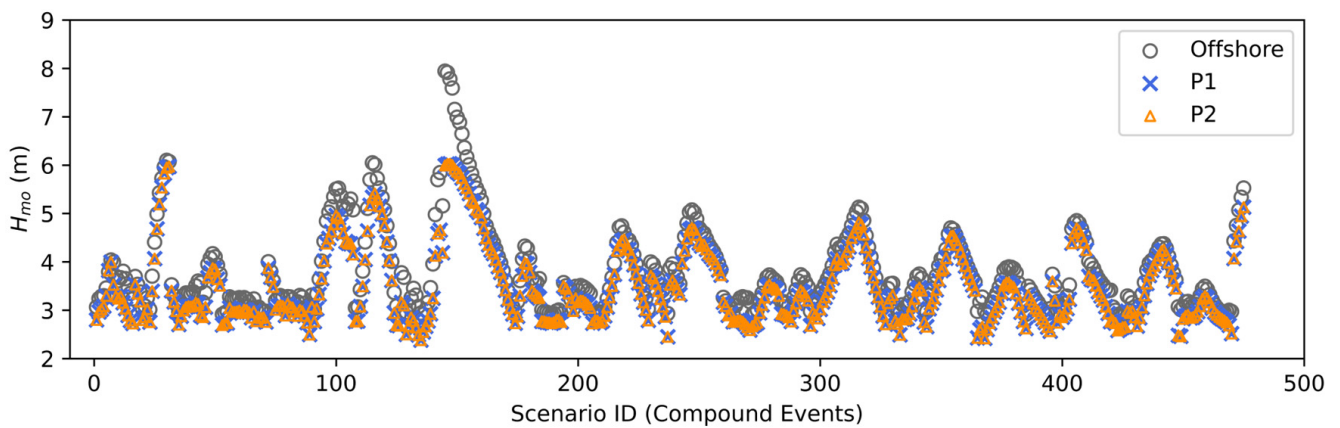
Figure 8. Comparison of significant wave heights for the historical dataset (1977–2005) between offshore wave data (circular markers), nearshore point P1 (x markers), and nearshore point P2 (triangular markers).
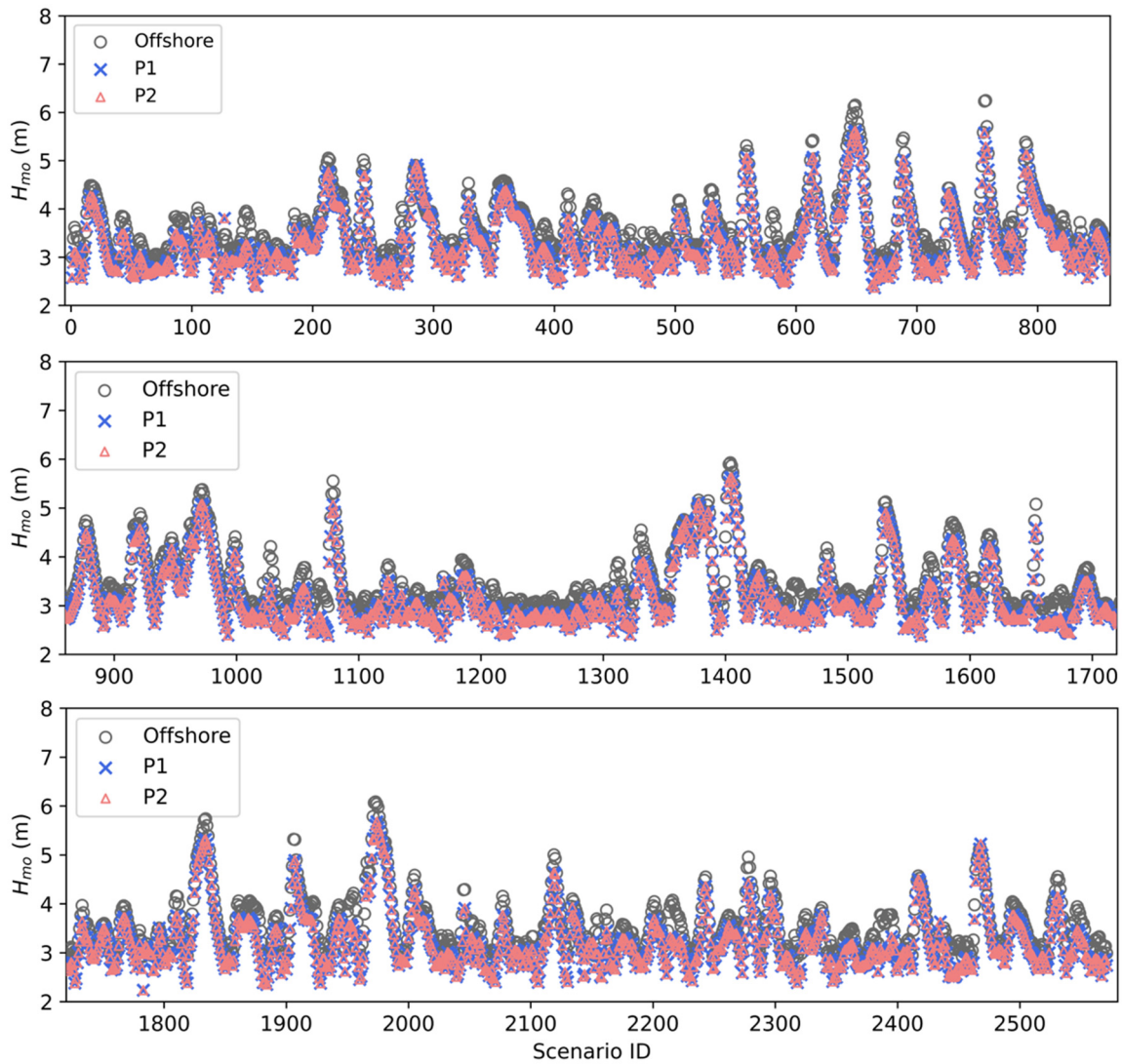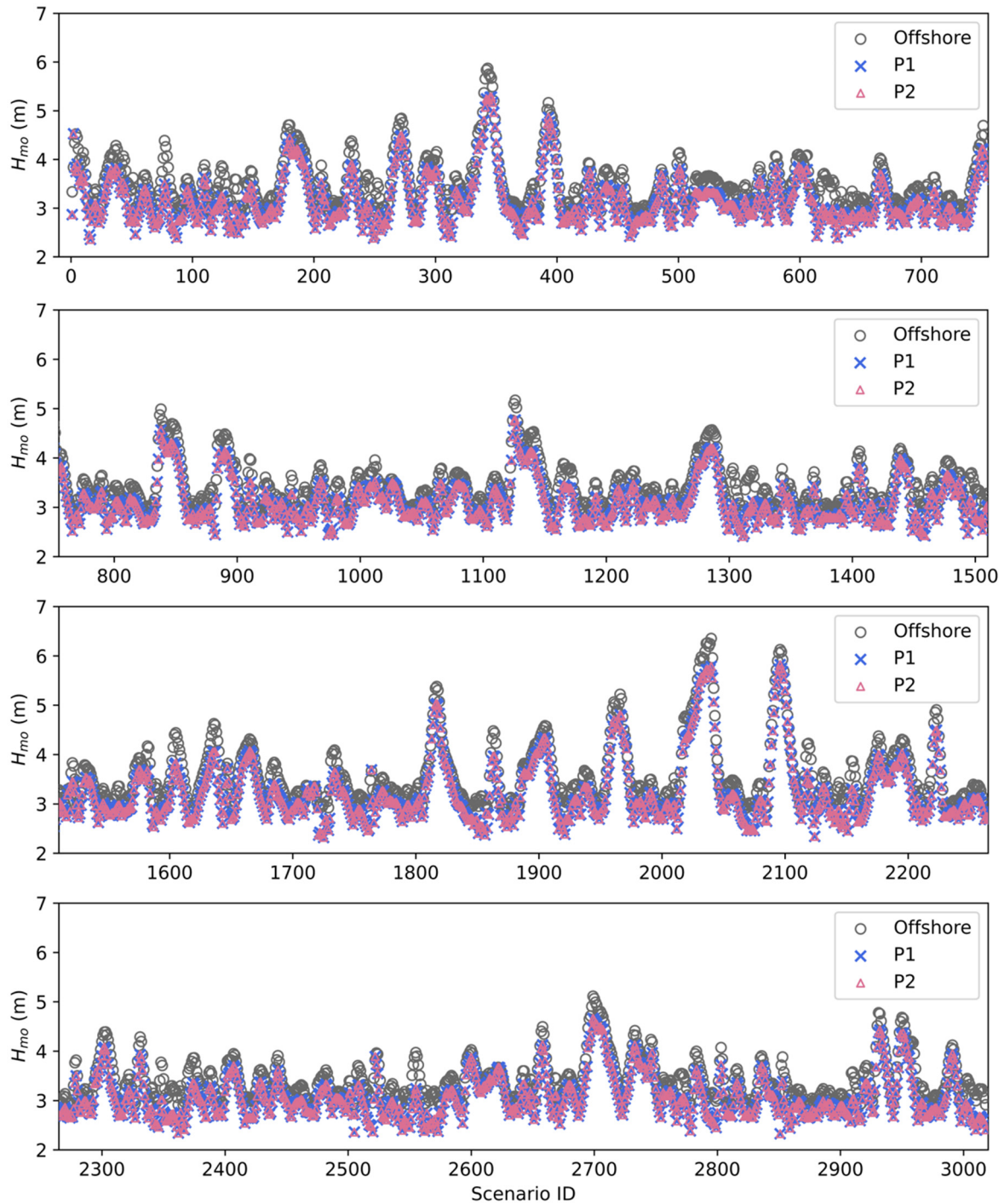
Figure 9. Comparison of significant wave heights for the RCP8.5 dataset (2041–2070) between offshore wave data (circular markers), nearshore point P1 (x markers), and nearshore point P2 (triangular markers). Vertical plots indicate sequence of scenarios.

As expected, a decrease in the wave height is observed from offshore to nearshore and is mainly attributed to the processes of wave refraction and shoaling. For instance, for the historical period an average decrease of 6% was observed, whereas the respective values for RCP 8.5 and RCP 4.5 were 6.8% and 7.33%. Both nearshore positions exhibit almost identical values of wave characteristics and so, considering that both coastal profiles are relatively similar up until the top of the sand dunes, only profile 2 will be used hereafter for the simulations using the CSHORE numerical model.

3.3. Wave Overtopping

Having calculated the nearshore wave characteristics, the next step concerns the calculation of wave overtopping for the compound events of the three examined datasets utilizing the CSHORE [46] coastal profile model. The main input data required by the model are the root mean square wave height H_{rms}, the peak wave period T_p, and the wave incidence angle a, all of which are provided by the obtained results of the PMS wave model. The model also takes into account the effect of sea level increase, and each wave condition is associated with the corresponding WL elevation, as obtained by the analysis of the compound events carried out in Section 3.1. It should be mentioned however that, being a coastal profile model, CSHORE has some inherent limitations as the full hydrodynamics are

not resolved and empirical relationships are utilized for the calculation of wave overtopping volumes, which may lead to inaccurate estimations, especially for steep beach profiles.


Figure 10. Comparison of significant wave heights for the RCP4.5 dataset (2071–2100) between offshore wave data (circular markers), nearshore point P1 (x markers), and nearshore point P2 (triangular markers). Vertical plots indicate sequence of scenarios.

Out of the 6232 simulated compound events, a total of 472 events lead to wave overtopping of the sand dunes, distributed as follows among each dataset: 24 in the historical dataset, 260 in RCP 8.5, and 188 in RCP 4.5. This corresponds to a percentage of wave overtopping of 5.0% for the historical, 10.1% for RCP 8.5, and 6.3% for RCP 4.5, signifying an increase in overtopping events over the coming years due to the effects of

climate change. It is noted that the morphological evolution of the coastal profile was not taken into account due to the unavailability of measurements that are required to calibrate the sediment transport parameters of the model.

Figure 11 depicts the relationship between H_{mo} and the other three examined parameters (i.e., WL, T_p, and a_o) for the combined datasets, distinguishing the events that lead to wave overtopping of the sand dunes.

(**a**)

(**b**)

(**c**)
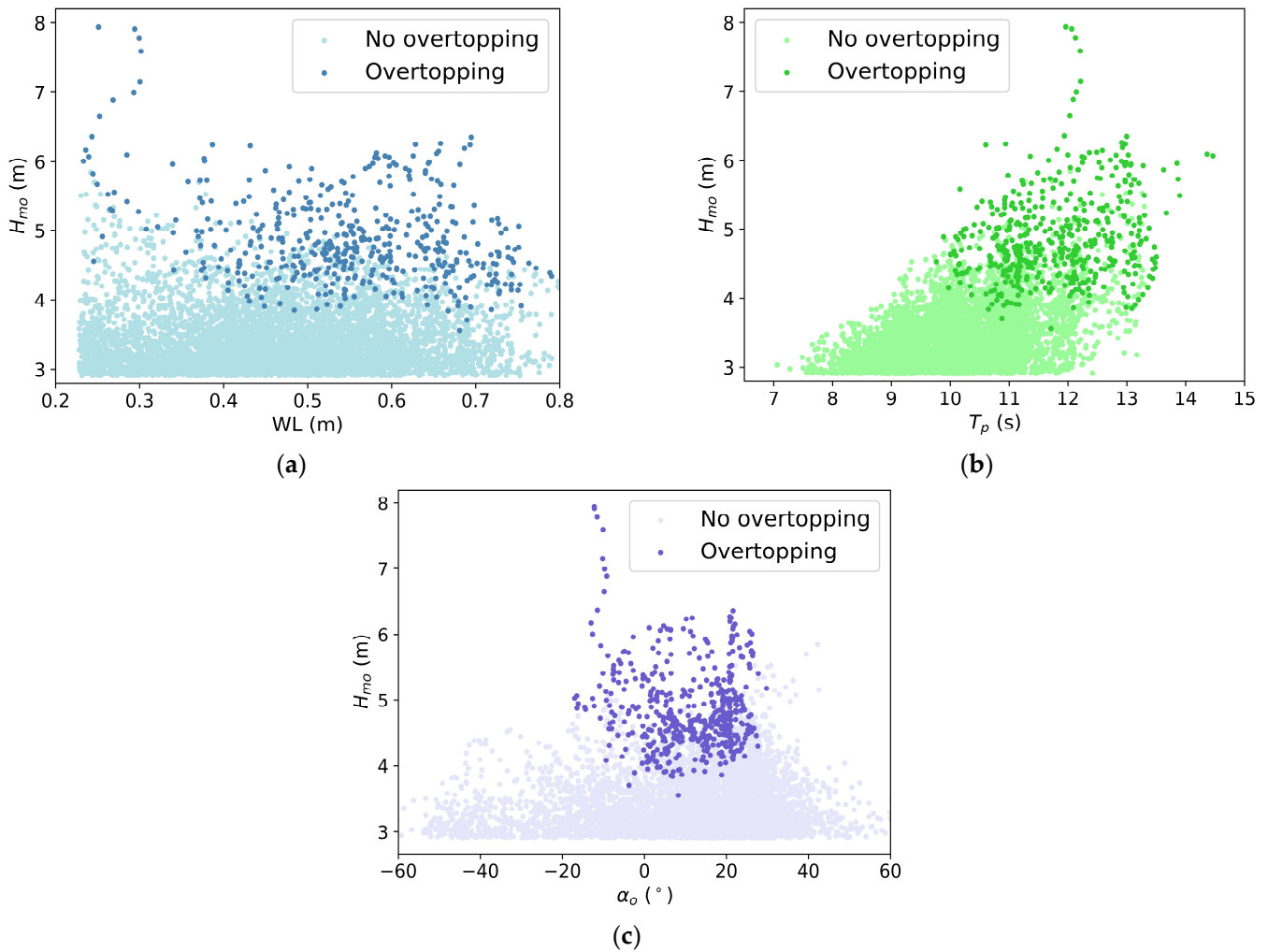
Figure 11. Compound spectral wave height events that generate wave overtopping in comparison with (**a**) total water level elevation, (**b**) peak wave period, and (**c**) deep water wave incidence angle.

While Figure 11 attempts to decompose wave overtopping and examine the individual relationship between the physical parameters examined, it should be mentioned that the interplay between them can potentially influence the occurrence of wave overtopping, Since wave overtopping is a complex physical process influenced by the aforementioned interactions, this is the potential reason why some highly energetic events in Figure 11 (as far as wave height is concerned) do not lead to wave overtopping.

3.4. Feature Weight Estimation

Apart from the analysis to estimate the interdependencies between the wave characteristics and total water level, an attempt is made to estimate the weight of each variable with respect to the overtopping rate.

To this end, feature weight (FW) estimation methods [48] have been employed extensively in the past years in order to enhance machine learning algorithms' performance, as well as to assess the relevance of each variable in contributing to the output pat-

tern. In the present research, the main interest lies in estimating which of the variables (i.e., H_{mo}, T_p, a_o, WL) contribute the most to wave overtopping and can potentially lead to coastal inundation, or if there is a variable that can be omitted from the analysis.

For the compound flood events identified for each of the three datasets (historical, RCP4.5, and RCP8.5), a Random Forest (RF) algorithm, commonly used for classification and regression tasks [70] was implemented. The algorithm was applied in the scikit-learn library [71], utilizing 100 estimators and a training/test dataset split of 90%/10%. This fractional split was selected over a more common split percentage of 80%/20% since the main focus is the feature weight estimation in the training dataset and not the predictive capabilities of the Random Forest classifier. The input variables of the Random Forest are normalized values of H_{mo}, T_p, a_o, and WL and the output is the overtopping rate, represented by a logical constructor with a "true" value if overtopping occurs and "false" if the particular combination of the input variables does not lead to wave overtopping. To estimate the feature importances of each of the input variables, a permutation importance technique was utilized. This technique is particularly useful for nonlinear or opaque estimators, and revolves around randomly shuffling the values of a single feature and observing the resulting degradation of the model's score [70]. It was selected over an impurity-based feature importance method since the latter can lead to misleading results for features with many unique values (high cardinality). The obtained mean accuracy decreases for each feature with respect to the examined dataset are illustrated in Figure 12. It is noted that a higher value indicates the increased importance of that feature.

As can be observed in Figure 12, negative feature importances are not identified, signifying that all variables contribute to the calculation of wave overtopping and none should be omitted from the analysis. This is a particularly interesting finding for the subsequent cluster analysis, since it reaffirms that all examined variables should be utilized as input for the clustering algorithm. Comparing the estimated weights between the datasets, it can be observed that the largest contribution to the wave overtopping event is predominantly from the spectral wave height followed by the peak wave period. Lower contributions are observed for the total water level and the offshore wave incidence angle throughout all examined datasets. As expected, the contribution of the total water level in wave overtopping increases for the RCP datasets. Specifically, a mean accuracy decrease of 3.6% for the historical dataset if water level is excluded becomes 7.8% for RCP 8.5 and 8.8% for RCP 4.5.

The weights obtained by the feature weight estimation analysis are subsequently provided through a function as a sample weight to a clustering algorithm with the ultimate goal of shifting the cluster centroids closer to combinations of input variables that are expected to lead to wave overtopping. The function selected is based on a simple weighted sum model [49] calculated for each combination of input variables as follows:

$$WSM = WL_{norm} \cdot w_{WL} + H_{mo,norm} \cdot w_{Hmo} + T_{p,norm} \cdot w_{Tp} + \cos(a_{o,norm}) \\ \cdot w_{ao} + fl_{overtopping} \tag{6}$$

where w denotes the individual weights for each variable, as dictated by the feature weight estimation, and $fl_{overtopping}$ is an integer receiving discrete values: 1 if the compound flood event leads to wave overtopping and 0 if it does not. The subscript "*norm*" denotes normalized values of each variable, obtained by implementing a min–max normalization method.

It is noted that the cosine of the offshore wave incidence was selected as a variable to calculate *WSM*, as a way to include the obliqueness of the waves, which leads to smaller overtopping volumes, as shown in Figure 11c.
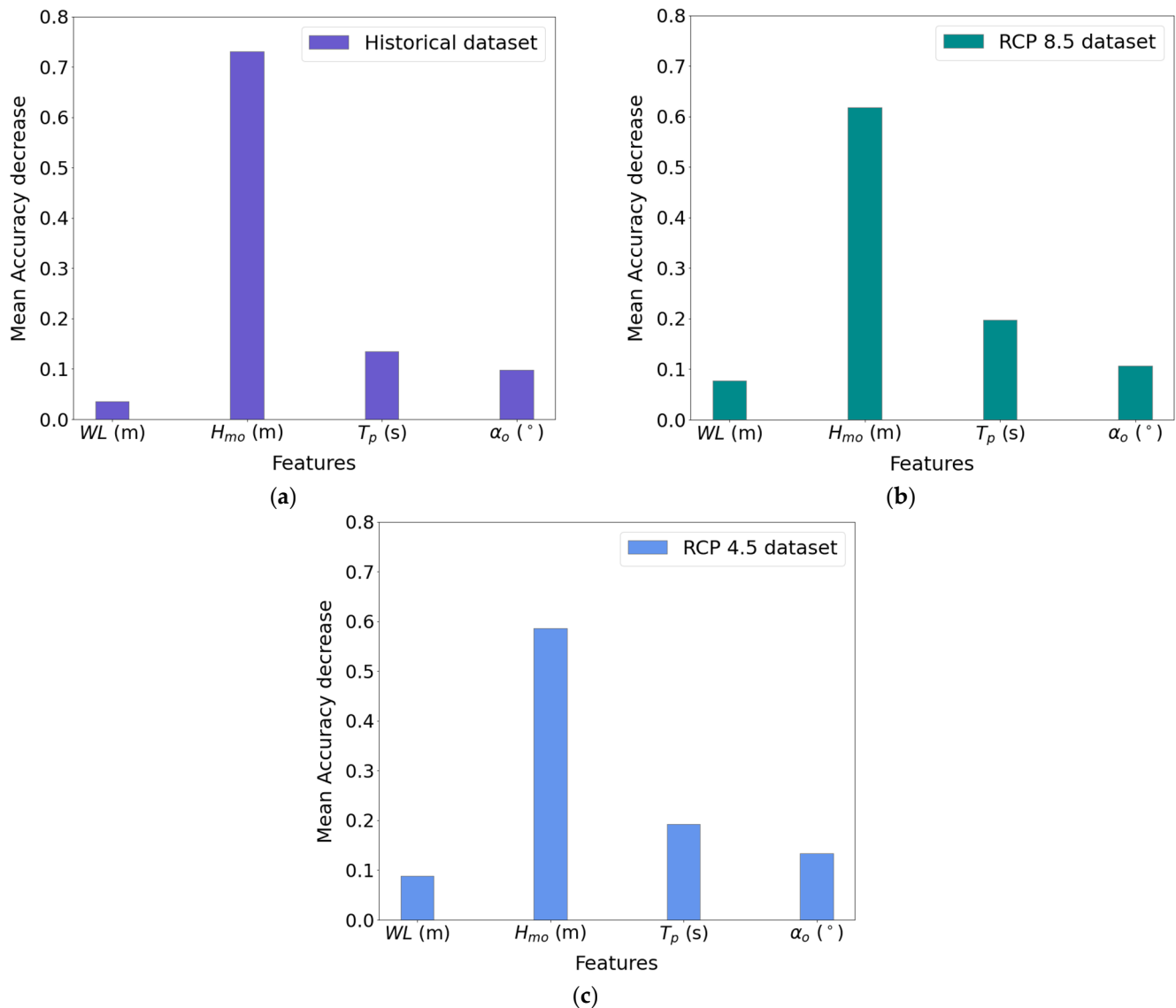
Figure 12. Bar plot of estimated feature importances for (**a**) historical dataset, (**b**) RCP 8.5, and (**c**) RCP 4.5.

3.5. Implementation of Clustering Analysis

The final step of the methodology concerns conducting clustering analysis in order to define the individual events that will be used to conduct hydrodynamic simulations, and thereafter, train a machine learning algorithm to be used in the framework of a coastal flooding EWS.

As mentioned in Section 3.1, each examined dataset was divided in two subcategories: a "supercritical" one, containing the compound flood events; and a "subcritical" one, containing the combinations of variables below the thresholds, categorized as non-compound events.

To carry out the clustering analysis, the well-known centroid-based K-Means clustering algorithm [53] was implemented. The reason for this selection over other alternatives is the capability of the K-Means algorithm to handle large datasets and the property to preemptively select the optimal number of clusters. The latter is especially important as it allows a degree of freedom to the number of obtained centroids.

An instance of the K-Means clustering algorithm was applied with the initial centroids selected through the K-Means++ algorithm [72]. For the number of iterations, 800 was set

as the optimal number to avoid fluctuations of the obtained centroids over consecutive implementations of the algorithm and maintain reasonable computational times.

The input variables of the cluster analysis were H_{mo}, T_p, $\cos a_o$, and WL which were normalized prior to inclusion in the clustering algorithm. It should be stated that for the subcritical dataset additional thresholds were imposed prior to the clustering algorithm. Therefore, all events with $H_{mo} < 2$ m and $T_p < 8.25$ s were excluded from the clustering analysis. The reason thresholds were imposed solely on these two variables is due to their increased importance in leading to wave overtopping. The value of the wave period was correlated with the threshold wave height and was estimated by performing regression for the sea states of all the datasets associated with $H_{mo} > 2$ m based on a power law [73]. Even though, from the compound event assessment it was discovered that events with $H_{mo} < 4$ m did not lead to wave overtopping, a lower threshold value for the wave heights was selected to provide a more diverse set of conditions to develop the EWS.

Importantly, it is mentioned that each combination of the input variables was accompanied by the calculated value of WSM, which was used as a sample weight in the instance of the K-Means algorithm. It is noted that in the case that the combination of variables examined is not a compound flood event, the value of $fl_{overtopping}$ in Equation (6) is set to zero.

The first point of the conducted clustering analysis was to define the optimal number of clusters and centroids to represent the subcritical and supercritical datasets. The main goal of this investigation was the selection of the appropriate number of centroids that would be, on one hand, able to prescribe in good detail the features of the full dataset but, on the other hand, to avoid prescribing an overly large amount of centroids which would unequivocally lead to several simulation scenarios to train the EWS. To select the optimal number of clusters and avoid overfitting, the well-known elbow method [74], using the distortion score as a metric, was utilized. The elbow method is a graphical approach which operates by calculating the within-cluster sum of squares (WCSS), which is the total of the squared distances between data points and their cluster center (k). The elbow is defined as the point where increasing k no longer leads to a significant decrease in the WCSS, and the rate of decrease slows down. Firstly, the elbow point was calculated for the subcritical historical dataset by incrementally increasing the number of clusters within the range of 10–150. The graphical representation of the elbow graph is shown in Figure 13. As can be seen from Figure 13, the elbow point of the graph was located at a number of $k = 50$ clusters.
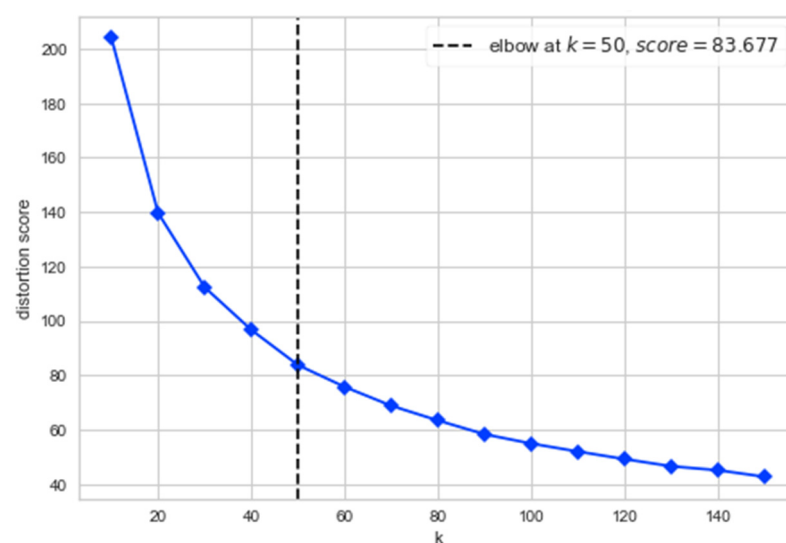

Figure 13. Elbow graph for the subcritical historical dataset.

It should be noted that the elbow point could be taken at a slightly higher number of clusters (e.g., 60 or 80); however, $k = 50$ is considered an adequate number to prescribe the subcritical historical dataset. Having selected the optimal number of clusters, Figure 14 shows the formation of clusters and corresponding centroids for two distinct cases: (a) without the utilization of sample weighting function in the K-Means algorithm, and (b) using the WSM value of each observation as a weight.

Plotting the clusters as a 2D scatter graph of H_{mo} against WL leads to a cluster formation with overlapping members, as is expected since the dataset contains four variables used as input in the clustering analysis. As can be deduced from Figure 14, the weighting function shifts the centroids to sea states that lie closer to the defined wave height and total water level threshold. This behavior is the intended one, as it is desirable to include combinations of sea states and water level that could potentially lead to overtopping in the training of the EWS. Consequently, the inclusion of weights in the analysis shifts the centroids to more extreme compound events, which is particularly important in the context of the development of an EWS framework for the prediction of coastal flooding.

For the supercritical historical dataset, containing the compound flood events, the determination of the number of clusters was based on the discretion of the authors instead of utilizing the elbow method because it is desirable to represent the compound events present in this dataset in much more detail, since a select few of them lead to wave overtopping of the sand dunes and coastal inundation. Consequently, it was chosen to utilize at least double the number of cluster centroids of the respective subcritical dataset. It should also be highlighted that utilizing the WSM quantity for the sample weight in the K-Means algorithm, calculated through Equation (6), ensures, on one hand, that the cluster centroids are shifted to more energetic sea states and higher total water levels, but also, that compound flood events that lead to wave overtopping are considered in the analysis with increased importance. Figure 15 depicts the cluster centroids for the historical supercritical dataset as a 2D scatterplot, superimposed with all the compound events. The dark blue markers denote events that lead to wave overtopping of the sand dunes.

As can be seen from Figure 15, the cluster centroids are well formed, and importantly, a select few of them are positioned close to events that lead to wave overtopping.

Using the same principles as the historical dataset, the configured K-Means algorithm was applied to both RCP 4.5 and RCP 8.5, obtaining 50 clusters for the subcritical and 100 clusters for the supercritical datasets, respectively. To illustrate the effect of incorporating weights in the clustering algorithm, Figure 16 shows the obtained clusters for the supercritical dataset of RCP 8.5 with (red triangular markers) and without (orange square markers) the implementation of the weighting function.

Although at first glance the differences between some of the obtained cluster centroids seem inconspicuous, a closer look reveals that incorporating weights in the K-Means algorithm can lead to some notable differences. In general, the centroids obtained by the K-Means algorithm incorporating weights are shifted towards higher values of wave height and total water level. Interestingly, for individual events that lead to wave overtopping scattered among others that do not overtop the sand dunes, the weighting function shifts the centroids towards these events. One such centroid is located at values of $H_s = 4.6$ m and $WL = 0.72$ m, while the respective cluster centroid without the inclusion of weights would lie in an area with events that are considered unable to overtop the sand dunes. Consequently, utilizing weights in the clustering analysis is expected to have more visible effects the larger the size of the dataset and the more overtopping events are observed, so the weights of these observations have a larger contribution.
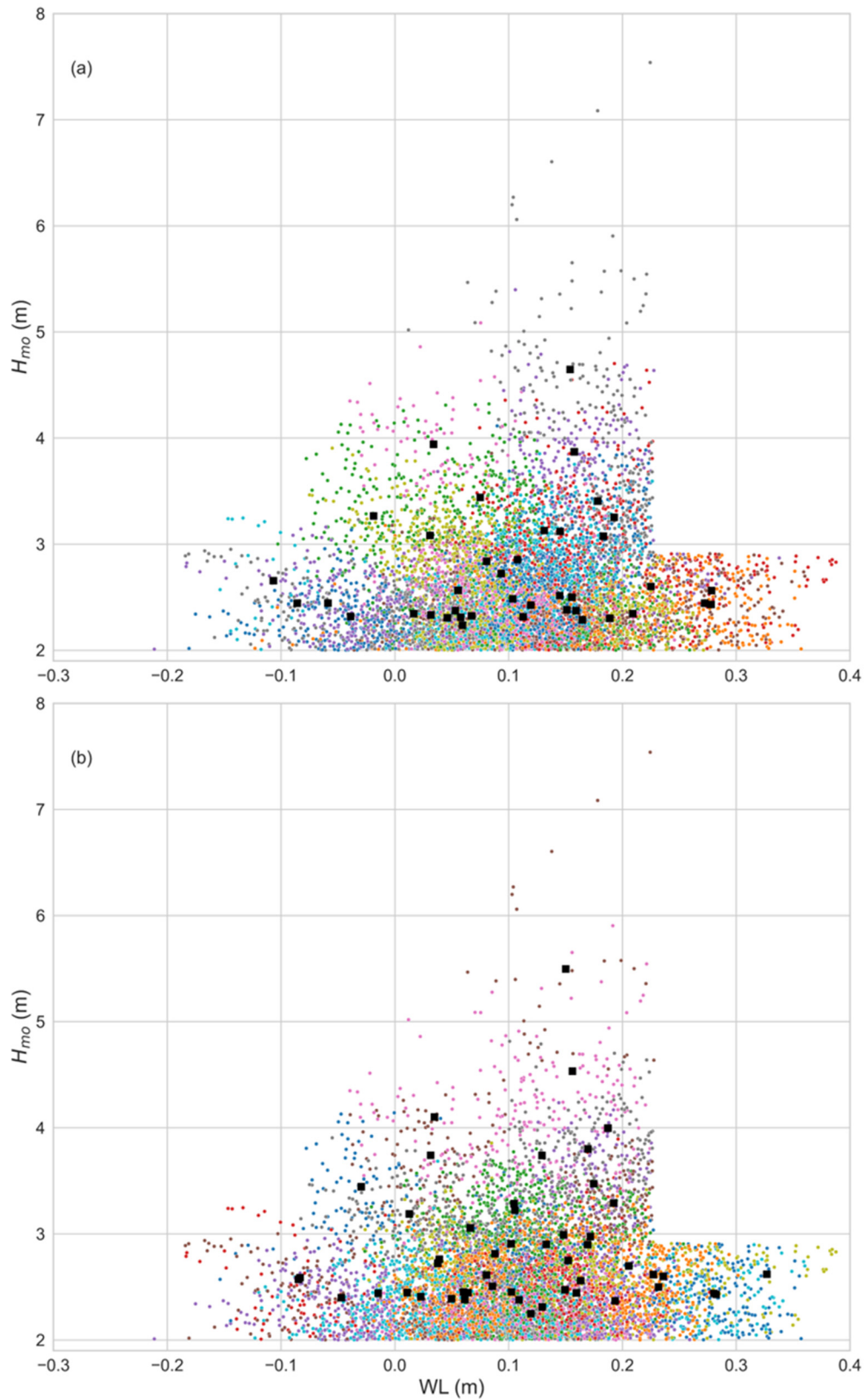
Figure 14. Obtained clusters and centroids for the subcritical historical dataset using (**a**) no sample weighting function and (**b**) the sample weighting function based on permutation feature importance. The different colors of the points indicate the cluster membership.
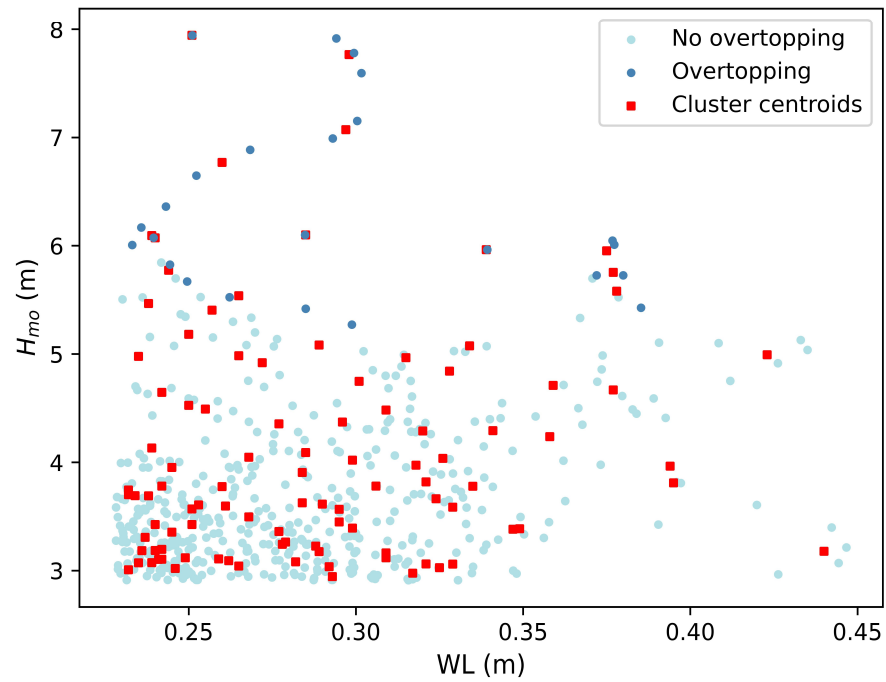
Figure 15. Compound events that generate wave overtopping superimposed with the centroids of the cluster analysis for the historical dataset.
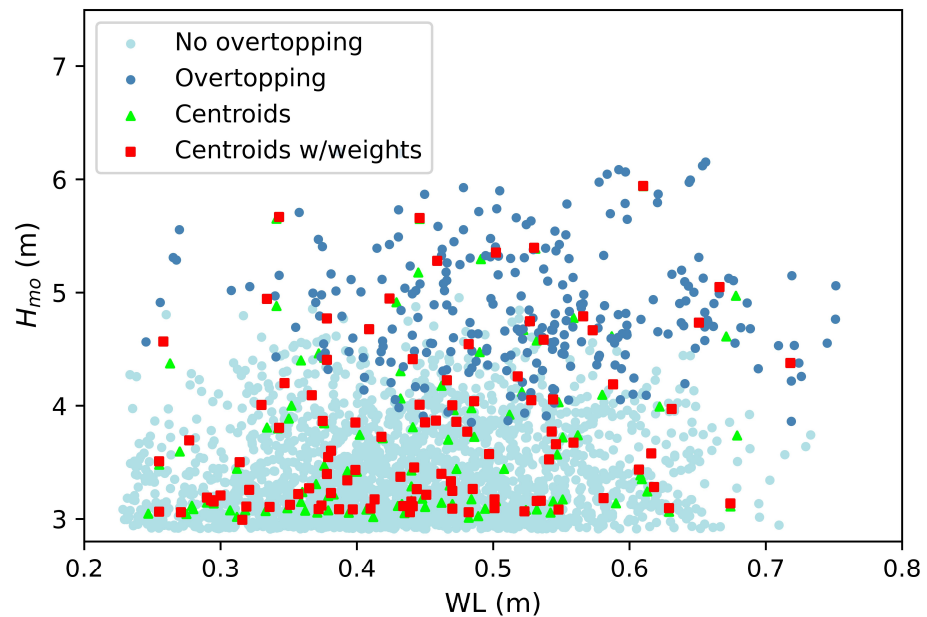
Figure 16. Compound events that generate wave overtopping superimposed with the centroids of the cluster analysis for the RCP 8.5 dataset without weighting function (green triangular markers) and with weighting function (red square markers).

In summary, by implementing the K-Means algorithm in each dataset, a total of 450 combinations are distinguished (150 for each dataset) and correspond to the obtained centroids from the clustering analysis. The obtained clusters for each dataset are compiled and shown as a 3D scatter plot in Figure 17 for the variables H_{mo}, T_p, and $\cos(a_o)$. It is considered that the computed centroids adequately represent the full dataset, but significantly reduce the large number of simulations that have to be conducted, and subsequently used, to develop a machine learning-based EWS for coastal flooding.
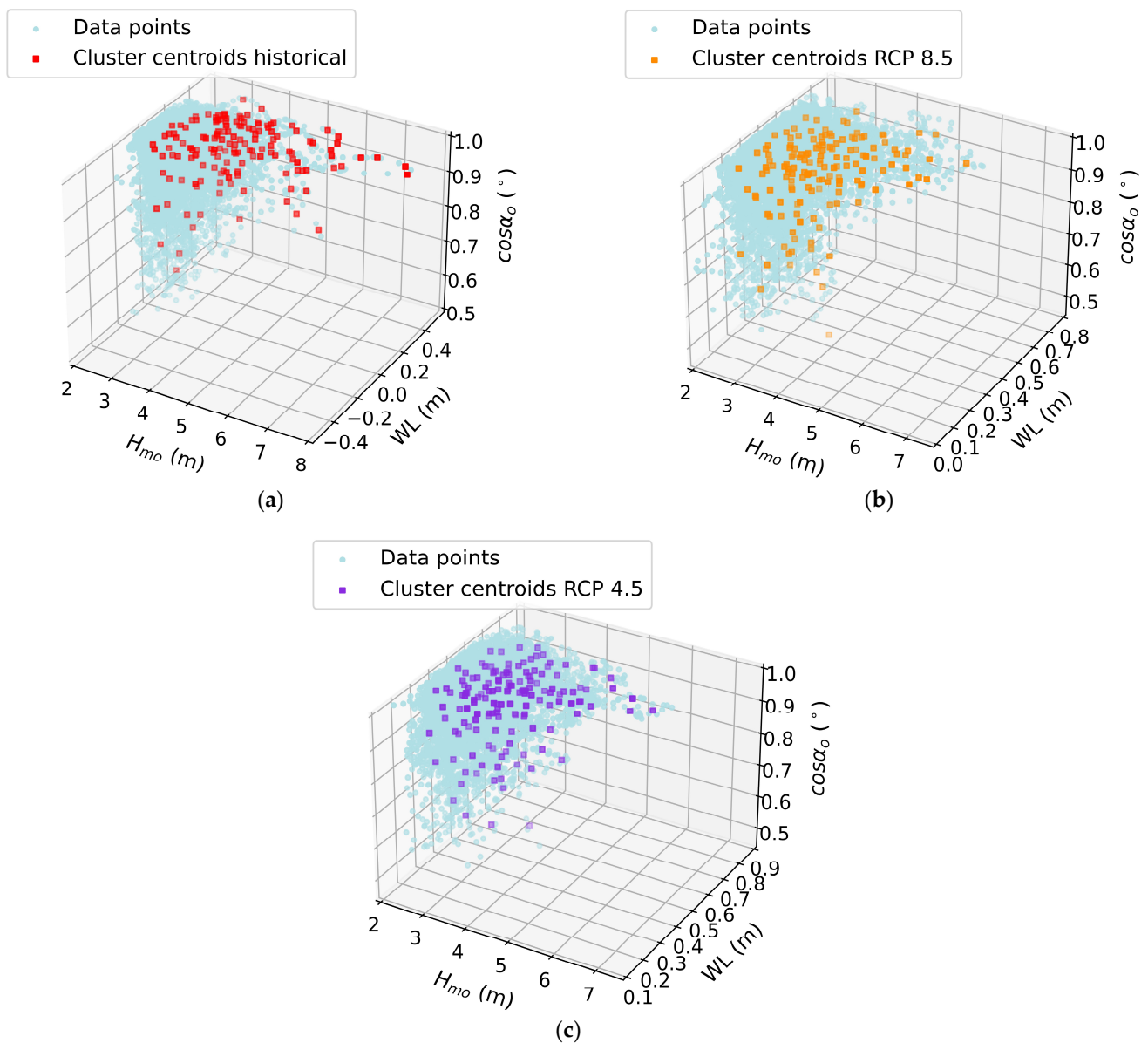
Figure 17. Obtained clusters for each dataset: (**a**) historical, (**b**) RCP 8.5, and (**c**) RCP 4.5.

3.6. Suggestions for Future Research

The proposed methodology presents an integrated framework to select representative sea states with the associated water level elevations, taking into account climate change projections, for applications in coastal flooding EWSs. The steps of the proposed methodology were also conceptualized in a manner allowing for many future research objectives to further expand and verify the proposed methodology.

An especially interesting topic for future research revolves around the inclusion of sediment transport and erosion of the coastal profile in the CSHORE model. Due to the inherent complexity in accurately modeling coastal processes, topographic surveys of the coastal profile, ideally before and after storm events, should take place in order to calibrate the parameters of CSHORE. However, the inclusion of the morphological evolution of the coastal profile can exacerbate the risk of coastal flooding due to wave overtopping, especially if simulations show that the sand dunes in the study area are in danger of eroding, compromising the flood protection of the hinterland.

Another objective for future research revolves around the evaluation of the performance of the machine learning algorithm that will be tasked with predicting the flooding

inundation in the framework of the EWS. It would be interesting to intercompare the performance of two EWSs, one based on simulation scenarios with the K-Means algorithm without the proposed weighting function, and another with the centroids calculated with the inclusion of weights in the algorithmic procedure. This comparison would further support the notion that the second EWS would be better equipped to predict events that lead to coastal flooding. This comparison could also enable the utilization of alternative clustering algorithms instead of K-Means, or examine if a hybrid approach combining the use of K-Means and MDA could ultimately enhance the performance of the EWS.

4. Conclusions

This research presents a robust methodology to predict coastal flooding through a framework that combines numerical modeling of nearshore wave propagation and clustering algorithms to select compound events that can potentially lead to inundation of the hinterland. The developed methodology, which was applied to the coastal area surrounding the estuary of Alfios river in Pyrgos, Greece, is based on processing of time series of wave characteristics and total water level elevation. The data refer to three distinct time periods: a historical period (1977–2005); the medium-term future (2041–2070), based on climate projections with RCP 8.5; and the long-term future (2071–2100), based on the RCP 4.5 emission scenario.

The analysis shows a significant increase in the frequency of compound flood events in future climate scenarios (RCP 4.5 and RCP 8.5) compared to the historical period. This indicates a higher risk of extreme events, which highlights the importance of developing a robust EWS to alert coastal communities to the imminent threat of coastal flooding.

In an integral step for the proposed methodology, feature weight estimation using a Random Forest algorithm was carried out to assess the contribution of each variable in wave overtopping. The conducted analysis indicates that, out of all the variables, spectral wave height and peak wave period are the most critical variables influencing wave overtopping. This finding is consistent across all datasets, with water level and wave incidence angle having lower contributions. Importantly, the contribution of water level elevation increases for the future scenarios incorporating climate change.

The computed weights from the future weight estimation are used to form a weighting function that is supplied as input to the K-Means clustering algorithm, that is tasked with selecting the representative scenarios for the development of an EWS for coastal flooding. Importantly, introducing a weighted sum model ensures that the most critical events are prioritized, shifting the centroids of the cluster analysis to the combination of more energetic sea states and higher values of total water level elevation. It is suggested for hourly time series of offshore wave characteristics to select at least 150 centroids (100 for the compound events and 50 for the rest of the dataset) to ensure a good description of the prevalent meteorological and oceanographic conditions of the study area.

While the proposed methodology offers significant advantages, such as its reproducibility and ability to prioritize critical compound flood events, there are some limitations that should be acknowledged. The methodology relies on certain assumptions in the numerical modeling and clustering processes, which could introduce uncertainties, particularly in extreme future scenarios, where the dynamics may not be fully captured by the models. Additionally, the clustering process, while effective and robust, may not perfectly reflect the full range of complex coastal interactions and dataset variability, particularly in highly dynamic environments.

In summary, the proposed methodology lays the groundwork for the optimal selection of scenarios in the context of coastal flooding early warning systems, being simultaneously easily reproducible, so it can be applied in any coastal area without restrictions, granted that the intricacies and unique features of each coastal area are properly represented. This approach has strong implications for enhancing climate resilience in coastal communities and informing coastal management policies by providing a data-driven foundation for proactive flood risk mitigation strategies.

References

1. Vousdoukas, M.I.; Mentaschi, L.; Voukouvalas, E.; Bianchi, A.; Dottori, F.; Feyen, L. Climatic and socioeconomic controls of future coastal flood risk in Europe. *Nat. Clim. Chang.* **2018**, *8*, 776–780. [CrossRef]
2. Calvin, K.; Dasgupta, D.; Krinner, G.; Mukherji, A.; Thorne, P.W.; Trisos, C.; Romero, J.; Aldunce, P.; Barrett, K.; Blanco, G.; et al. *IPCC, 2023: Climate Change: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change: Geneva, Switzerland, 13–19 March 2023*; Lee, H., Romero, J., Eds.; IPCC: Geneva, Switzerland, 2023. [CrossRef]
3. Wang, Y.; Gao, X.; Wei, J.; Deng, Y.; Yang, X. Projections of Coastal Flooding under Different RCP Scenarios over the 21st Century: A Case Study of China's Coastal Zone. *Estuar. Coast. Shelf Sci.* **2023**, *282*, 108155. [CrossRef]
4. Roy, P.; Pal, S.C.; Chakrabortty, R.; Chowdhuri, I.; Saha, A.; Shit, M. Effects of Climate Change and Sea-Level Rise on Coastal Habitat: Vulnerability Assessment, Adaptation Strategies, and Policy Recommendations. *J. Environ. Manag.* **2023**, *325*, 117187. [CrossRef] [PubMed]
5. Wang, S.; Najafi, M.R.; Cannon, A.J.; Khan, A.A. Uncertainties in Riverine and Coastal Flood Impacts under Climate Change. *Water* **2021**, *13*, 1774. [CrossRef]
6. Heinrich, P.; Hagemann, S.; Weisse, R.; Schrum, C.; Daewel, U.; Gaslikova, L. Compound flood events: Analysing the joint occurrence of extreme river discharge events and storm surges in northern and central Europe. *Nat. Hazards Earth Syst. Sci.* **2023**, *23*, 1967–1985. [CrossRef]
7. Heinrich, P.; Hagemann, S.; Weisse, R.; Gaslikova, L. Changes in compound flood event frequency in northern and central Europe under climate change. *Front. Clim.* **2023**, *5*, 1227613. [CrossRef]
8. Paprotny, D.; Kreibich, H.; Morales-Nápoles, O.; Castellarin, A.; Carisi, F.; Schröter, K. Exposure and vulnerability estimation for modelling flood losses to commercial assets in Europe. *Sci. Total Environ.* **2020**, *737*, 140011. [CrossRef]
9. Malliouri, D.I.; Kyriakidou, C.; Moraitis, V.; Vandarakis, D.; Martzikos, N.; Gad, F.K.; Hatiris, G.A.; Kapsimalis, V. A new approach for the assessment of coastal flooding risk. Application in Rhodes island, Greece. *Appl. Ocean Res.* **2024**, *148*, 104006. [CrossRef]
10. Tsaimou, C.N.; Papadimitriou, A.; Chalastani, V.; Sartampakos, P.; Chondros, M.; Tsoukala, V.K. Impact of Spatial Segmentation on the Assessment of Coastal Vulnerability—Insights and Practical Recommendations. *J. Mar. Sci. Eng.* **2023**, *11*, 1675. [CrossRef]
11. Shen, P.; Wei, S.; Shi, H.; Gao, L.; Zhou, W.H. Coastal Flood Risk and Smart Resilience Evaluation under a Changing Climate. *Ocean. Res.* **2023**, *2*, 0029. [CrossRef]
12. Rogers, D.; Tsirkunov, V. *Global Assessment Report on Disaster Risk Reduction: Costs and Benefits of Early Warning Systems*; The World Bank: Washington, DC, USA, 2010.
13. Harley, M.D.; Valentini, A.; Armaroli, C.; Perini, L.; Calabrese, L.; Ciavola, P. Can an early-warning system help minimize the impacts of coastal storms? A case study of the 2012 Halloween storm, northern Italy. *Nat. Hazards Earth Syst. Sci.* **2016**, *16*, 209–222. [CrossRef]
14. Espejo, A.; Wandres, M.; Damlamian, H.; Divesh, A.; Giblin, J.; Aucan, J.; Eria, M.; Toorua, U. Efficient coastal inundation early-warning system for low-lying atolls, dealing with lagoon and ocean side inundation in Tarawa, Kiribati. *Weather Clim. Extrem.* **2023**, *42*, 100615. [CrossRef]

15. Doong, D.J.; Chuang, L.Z.H.; Wu, L.C.; Fan, Y.M.; Kao, C.C.; Wang, J.H. Development of an operational coastal flooding early warning system. *Nat. Hazards Earth Syst. Sci.* **2012**, *12*, 379–390. [CrossRef]

16. Chondros, M.; Tsoukala, V.; Metallinos, A.; Papadimitriou, A.; Memos, C. A coastal flood early-warning system based on offshore sea state forecasts and artificial neural networks. *J. Mar. Sci. Eng.* **2021**, *9*, 1272. [CrossRef]

17. Garzon, J.L.; Ferreira; Zózimo, A.C.; Fortes, C.J.E.M.; Ferreira, A.M.; Pinheiro, L.V.; Reis, M.T. Development of a Bayesian networks-based early warning system for wave-induced flooding. *Int. J. Disaster Risk Reduct.* **2023**, *96*, 103931. [CrossRef]

18. Memarian Sorkhabi, O.; Shadmanfar, B.; Al-Amidi, M.M. Deep learning of sea-level variability and flood for coastal city resilience. *City Environ. Interact.* **2023**, *17*, 100098. [CrossRef]

19. Coppernicus Marine Environment Monitoring Service CMEMS. Available online: http://marine.copernicus.eu/ (accessed on 26 July 2024).

20. ECMWF. Available online: https://www.ecmwf.int/ (accessed on 25 July 2024).

21. NOAA. Available online: https://www.ncdc.noaa.gov/data-access (accessed on 25 July 2024).

22. Benedet, L.; Dobrochinski, J.P.F.; Walstra, D.J.R.; Klein, A.H.F.; Ranasinghe, R. A morphological modeling study to compare different methods of wave climate schematization and evaluate strategies to reduce erosion losses from a beach nourishment project. *Coast. Eng.* **2016**, *112*, 69–86. [CrossRef]

23. Papadimitriou, A.; Panagopoulos, L.; Chondros, M.; Tsoukala, V. A wave input-reduction method incorporating initiation of sediment motion. *J. Mar. Sci. Eng.* **2020**, *8*, 597. [CrossRef]

24. Papadimitriou, A.; Chondros, M.; Metallinos, A.; Tsoukala, V. Accelerating Predictions of Morphological Bed Evolution by Combining Numerical Modelling and Artificial Neural Networks. *J. Mar. Sci. Eng.* **2022**, *10*, 1621. [CrossRef]

25. de Queiroz, B.; Scheel, F.; Caires, S.; Walstra, D.J.; Olij, D.; Yoo, J.; Reniers, A.; de Boer, W. Performance evaluation of wave input reduction techniques for modeling inter-annual sandbar dynamics. *J. Mar. Sci. Eng.* **2019**, *7*, 148. [CrossRef]

26. Papadimitriou, A.; Tsoukala, V. Evaluating and enhancing the performance of the K-Means clustering algorithm for annual coastal bed evolution applications. *Oceanologia* **2024**, *66*, 267–285. [CrossRef]

27. Chondros, M.K.; Metallinos, A.S.; Memos, C.D.; Karambas, T.V.; Papadimitriou, A.G. Concerted nonlinear mild-slope wave models for enhanced simulation of coastal processes. *Appl. Math. Model.* **2021**, *91*, 508–529. [CrossRef]

28. USACE. *HEC-RAS Hydraulic Reference Manual*; US Army Corps of Engineers: Davis, CA, USA, 2021.

29. Booij, N.; Ris, R.C.; Holthuijsen, L.H. A third-generation wave model for coastal regions 1. Model description and validation. *J. Geophys. Res. Ocean* **1999**, *104*, 7649–7666. [CrossRef]

30. Roelvink, D.; Reniers, A.; van Dongeren, A.; van Thiel de Vries, J.; McCall, R.; Lescinski, J. Modelling storm impacts on beaches, dunes and barrier islands. *Coast. Eng.* **2009**, *56*, 1133–1152. [CrossRef]

31. Camus, P.; Mendez, F.J.; Medina, R.; Cofiño, A.S. Analysis of clustering and selection algorithms for the study of multivariate wave climate. *Coast. Eng.* **2011**, *58*, 453–462. [CrossRef]

32. Blain, C.A.; Westerink, J.J.; Luettich, R.A.; Scheffner, N.W. *ADCIRC: An Advanced Three-Dimensional Circulation Model for Shelves Coasts and Estuaries, Report 4: Hurricane Storm Surge Modeling Using Large Domains*; Dredging Research Program Technical Report DRP-92-6; U.S. Army Engineers Waterways Experiment Station: Vicksburg, MS, USA, 1994.

33. National Cadastre of Greece. Available online: https://www.ktimatologio.gr/ (accessed on 10 August 2024).

34. Copernicus Climate Data Store (CDS). Available online: https://cds.climate.copernicus.eu/#!/home (accessed on 10 August 2024).

35. Komen, G.J.; Cavaleri, L.; Donelan, M.; Hasselmann, K.; Hasselmann, S.; Janssen, P.A.E.M. *Dynamics and Modelling of Ocean Waves*; Cambridge University Press: Cambridge, UK, 1994.

36. Tsoukala, V.K.; Chondros, M.; Kapelonis, Z.G.; Martzikos, N.; Lykou, A.; Belibassakis, K.; Makropoulos, C. An integrated wave modelling framework for extreme and rare events for climate change in coastal areas—The case of Rethymno, Crete. *Oceanologia* **2016**, *58*, 71–89. [CrossRef]

37. Martzikos, N.T.; Prinos, P.E.; Memos, C.D.; Tsoukala, V.K. Statistical analysis of Mediterranean coastal storms. *Oceanologia* **2021**, *63*, 133–148. [CrossRef]

38. Petroliagkis, T.I.; Voukouvalas, E.; Disperati, J.; Bidlot, J. *Joint Probabilities of Storm Surge, Significant Wave Height and River Discharge Components of Coastal Flooding Events. Utilising Statistical Dependence Methodologies and Techniques*; Technical Report JRC100839; Joint Research Center (JCR): Luxembourg, 2016.

39. Marcos, M.; Rohmer, J.; Vousdoukas, M.I.; Mentaschi, L.; Le Cozannet, G.; Amores, A. Increased Extreme Coastal Water Levels Due to the Combined Action of Storm Surges and Wind Waves. *Geophys. Res. Lett.* **2019**, *46*, 4356–4364. [CrossRef]

40. Kraus, N.C.; Larson, M.; Wise, R.A. Depth of closure in beach-fill design. *Coast. Eng. Tech. Note* **1998**, *3*, 98.

41. Tolman, H.L. A Third-Generation Model for Wind Waves on Slowly Varying, Unsteady, and Inhomogeneous Depths and Currents. *J. Phys. Oceanogr.* **1991**, *201*, 782–797. [CrossRef]

42. Benoit, M.; Marcos, F.; Becq, F. TOMAWAC: A prediction model for offshore and nearshore storm waves. In Proceedings of the Congress of the International Association of Hydraulic Research, IAHR, San Francisco, CA, USA, 10–15 August 1997.

43. Karambas, T.V.; Samaras, A.G. An integrated numerical model for the design of coastal protection structures. *J. Mar. Sci. Eng.* **2017**, *5*, 50. [CrossRef]

44. Chondros, M.K.; Metallinos, A.S.; Papadimitriou, A.G. Enhanced Mild-Slope Wave Model with Parallel Implementation and Artificial Neural Network Support for Simulation of Wave Disturbance and Resonance in Ports. *J. Mar. Sci. Eng.* **2024**, *12*, 281. [CrossRef]

45. van Gent, M.R.A.; van den Boogaard, H.F.P.; Pozueta, B.; Medina, J.R. Neural network modelling of wave overtopping at coastal structures. *Coast. Eng.* **2007**, *54*, 586–593. [CrossRef]

46. Kobayashi, N. *Documentation of Cross-Shore Numerical Model CSHORE*; Research Report No. CACR-09-06; Center for Applied Coastal Research, University of Delaware: Newark, DE, USA, 2009.

47. Van der Meer, J.W.; Allsop, N.W.H.; Bruce, T.; De Rouck, J.; Kortenhaus, A.; Pullen, T.; Schüttrumpf, H.; Troch, P.; Zanuttigh, B. *EurOtop: Manual on Wave Overtopping of Sea Defences and Related Structures. An Overtopping Manual Largely Based on European Research, but for Worldwide Application*; Ghent University: Ghent, Belgium, 2018.

48. Niño-Adan, I.; Manjarres, D.; Landa-Torres, I.; Portillo, E. Feature weighting methods: A review. *Expert Syst. Appl.* **2021**, *184*, 115424. [CrossRef]

49. Triantaphyllou, E. Multi-Criteria Decision Making Methods. In *Multi-Criteria Decision Making Methods. A Comparative Study*; Applied Optimization; Springer: Boston, MA, USA, 2000; Volume 44, pp. 5–21. ISBN 978-1-4419-4838-0.

50. Martzikos, N.; Afentoulis, V.; Tsoukala, V. Storm clustering and classification for the port of Rethymno in Greece. *Water Util. J.* **2018**, *20*, 67–79.

51. Splinter, K.D.; Golshani, A.; Stuart, G.; Tomlinson, R. Spatial and Temporal Variability of Longshore Transport Along Gold Coast, Australia. *Coast. Eng. Proc.* **2011**, *1*, 95. [CrossRef]

52. Kelpšaitė-Rimkienė, L.; Parnell, K.E.; Žaromskis, R.; Kondrat, V. Cross-shore profile evolution after an extreme erosion event—Palanga, Lithuania. *J. Mar. Sci. Eng.* **2021**, *9*, 38. [CrossRef]

53. MacQueen, J. Some methods for classification and analysis of multivariate observations. In Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 1 January 1967.

54. Bezdek, J.C. *Pattern Recognition with Fuzzy Objective Function Algorithms*; Springer: Berlin/Heidelberg, Germany, 1981.

55. Guha, S.; Rastogi, R.; Shim, K. CURE: An efficient clustering algorithm for large databases. *SIGMOD Rec.* **1998**, *27*, 73–84. [CrossRef]

56. Kirby, J.T.; Dalrymple, R.A. A parabolic equation for the combined refraction diffraction of Stokes waves by mildly varying topography. *J. Fluid. Mech.* **1983**, *136*, 453–466. [CrossRef]

57. Kirby, J.T. A general wave equation for waves over rippled beds. *J. Fluid. Mech.* **1986**, *162*, 171–186. [CrossRef]

58. Kirby, J.T. On the gradual reflection of weakly nonlinear Stokes waves in regions with varying topography. *J. Fluid. Mech.* **1986**, *162*, 187–209. [CrossRef]

59. Battjes, J.A.; Janssen, J.P.F.M. Energy Loss and Set-Up Due To Breaking of Random Waves. *Proc. Coast. Eng. Conf.* **1979**, *1*, 569–587.

60. Putnam, J.A.; Johson, J.W. The dissipation of wave energy by bottom friction. *Eos Trans. Am. Geophys. Union.* **1949**, *30*, 67–74. [CrossRef]

61. Kirby, J.T.; Dalrymple, R.A. An approximate model for nonlinear dispersion in monochromatic wave propagation models. *Coast. Eng.* **1986**, *9*, 545–561. [CrossRef]

62. Kobayashi, N.; Hicks, B.S.; Figlus, J. Evolution of Gravel Beach Profiles. *J. Waterw. Port. Coast. Ocean Eng.* **2011**, *137*, 258–262. [CrossRef]

63. Kobayashi, N.; Farhadzadeh, A.; Melby, J.A. Wave Overtopping and Damage Progression of Stone Armor Layer. *J. Waterw. Port. Coast. Ocean. Eng.* **2010**, *136*, 257–265. [CrossRef]

64. Figlus, J.; Kobayashi, N.; Gralher, C.; Iranzo, V. Wave Overtopping and Overwash of Dunes. *J. Waterw. Port. Coast. Ocean. Eng.* **2011**, *137*, 26–33. [CrossRef]

65. Kobayashi, N.; Farhadzadeh, A. *Dune Erosion and Overwash*; World Scientific Pub Co Pte Ltd.: Singapore, 2009; pp. 1–14.

66. Battjes, J.A.; Stive, M.J.F. Calibration and verification of a dissipation model for random breaking waves. *J. Geophys. Res. Ocean.* **1985**, *90*, 9159–9167. [CrossRef]

67. Kobayashi, N.; Zhao, H.; Tega, Y. Suspended sand transport in surf zones. *J. Geophys. Res. Ocean.* **2005**, *110*, 1–21. [CrossRef]

68. Kobayashi, N.; Agarwal, A.; Johnson, B.D. Longshore Current and Sediment Transport on Beaches. *J. Waterw. Port. Coast. Ocean Eng.* **2007**, *133*, 296–304. [CrossRef]

69. Robins, P.; Lewis, M.; Elnahrawi, M.; Lyddon, C.; Dickson, N.; Coulthard, T. Compound Flooding: Dependence at Sub-daily Scales Between Extreme Storm Surge and Fluvial Flow. *Front. Built Environ.* **2021**, *7*, 727294. [CrossRef]

70. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

71. Pedregosa, F.; Michel, V.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Vanderplas, J.; Cournapeau, D.; Pedregosa, F.; Varoquaux, G.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

72. Arthur, D.; Vassilvitskii, S. K-means++: The advantages of careful seeding. In Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms, New Orleans, LA, USA, 7–9 January 2007.

73. Goda, Y. *Random Seas and Design of Maritime Structures*, 3rd ed.; Advanced Series on Ocean Engineering; World Scientific: Singapore, 2010; Volume 15, ISBN 9789814282413.
74. Joshi, K.D.; Nalwade, P.S. Modified K-Means for Better Initial Cluster Centres. *Int. J. Comput. Sci. Mob. Comput.* **2013**, *2*, 219–223.