

Article

Modelling and Clustering Sea Conditions: Bivariate Finite Mixtures of Generalized Additive Models for Location, Shape, and Scale Applied to the Analysis of Meteorological Tides and Wave Heights

Lorena Ricciotti ^{1,*} , Marco Picone ² , Alessio Pollice ¹  and Antonello Maruotti ³ ¹ Department of Economics and Finance, University of Bari Aldo Moro, 70121 Bari, Italy; alessio.pollice@uniba.it² ISPRA, 00144 Rome, Italy; marco.picone@isprambiente.it³ Department of Law, Economics, Politics and Modern Languages, LUMSA University, 00193 Rome, Italy; a.maruotti@lumsa.it

* Correspondence: lorena.ricciotti@uniba.it

Abstract: Modelling sea conditions is a complex task that requires a comprehensive analysis, considering various influencing factors. Observed and unobserved factors jointly play a role in the definition of sea conditions. Here, we consider finite mixtures of generalized linear additive models for location scale, and shape (GAMLSSs) to capture the effects of both environmental variables and omitted variables, whose effects are summarized using latent variables. The GAMLSS approach is flexible enough to allow for different data features such as non-normality, skewness, heavy tails, etc., and for the definition of a regression model not only for the expected values of the observed process but also for all the other distribution parameters, e.g., the variance. We collected data on multiple sea-related and environmental variables in Ancona (Italy) from two Italian networks: the Sea Level Measurement Network (Rete Mareografica Nazionale, RMN) and the Sea Waves Measurement Network (Rete Ondametrica Nazionale, RON). Our main outcomes were the meteorological tides (often also referred to as “residuals”) and the significant wave height. Atmospheric pressure and wind speed were considered as main drivers of the sea conditions, as well as the fetch associated with wind direction, linking these variables to the outcomes through the definition of multiple linear predictors in a regression framework. Our results confirm the importance of accounting for environmental variables and reveal that their effect is heterogeneous, where heterogeneity is modelled by three distinct mixture components, each capturing different sea conditions. These findings contribute to a deeper understanding of sea state dynamics and provide evidence of a clustering structure characterizing different sea conditions.

Keywords: meteorological tides; sea conditions; clustering; GAMLSS; finite mixtures; Adriatic Sea

Citation: Ricciotti, L.; Picone, M.; Pollice, A.; Maruotti, A. Modelling and Clustering Sea Conditions: Bivariate Finite Mixtures of Generalized Additive Models for Location, Shape, and Scale Applied to the Analysis of Meteorological Tides and Wave Heights. *J. Mar. Sci. Eng.* **2024**, *12*, 740. <https://doi.org/10.3390/jmse12050740>

Academic Editors: Angelo Rubino and Rafael J. Bergillos

Received: 11 March 2024

Revised: 23 April 2024

Accepted: 26 April 2024

Published: 29 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

When discussing the impact of climate change on the environment, it is crucial to monitor marine physical parameters as key indicators. Here, we focused on two outcomes, namely, meteorological tides (or residuals) and significant wave height, aiming at quantifying the impact of atmospheric pressure, wind, and effective fetch on the outcomes' expected value and variance. The crucial role of these variables is widely acknowledged in the literature. Atmospheric pressure plays a primary role, with a low pressure leading to high sea levels and vice versa. Wind speed and direction, as well as the fetch, are closely interconnected: wind speed can either decrease or increase wave heights depending on its direction, with the latter linked to the fetch, i.e., the stretch of the open sea over which the wind blows, determining wave generation. A greater fetch allows the wind to act upon a larger expanse of the sea surface, resulting in higher waves.

Exploring these relationships can be approached through various methodologies such as deterministic and statistical methods. In most cases, a deterministic approach is preferred by specialists for its ability to investigate large-scale variations and its aptitude to downscale to local scale. Nevertheless, the complexity of ocean phenomena, the variety of superposed processes, and the random nature of most of them make this approach prohibitive. This is more evident in particular geographical situations, such as semi-enclosed basins, with complex coastal orography, or with a high variability in met-ocean parameters at the air–water interface. In this framework, a statistical approach can help to capture dynamics that deterministic models are not able to depict as formulas in a closed form. In the last years, statistical models have been developed and improved, even coupled with deterministic models, to better investigate the marine environment.

Regrettably, real-world data often suffer from contamination by outliers, spurious points, or noise, generally referred to as bad points [1,2], which can adversely affect parameter estimates (see [3,4] for an overview of robust methods). However, as discussed by Davies and Gather [5], Hennig [6], it is of fundamental importance to remark that defining bad observations is relative to a reference distribution. Thus, the region of bad points can be delineated, for instance, by a region where the density of the reference distribution is low. Following other distribution-based approaches, such as those based on the Gaussian distribution, i.e., linear regression with Gaussian-distributed error terms, we opted for the t distribution as the reference distribution to account for heavy tails [7]. Thus, we considered a general class of univariate regression models, called generalized additive models for location, scale, and shape (GAMLSSs, [8–10]), where the exponential family assumption (as in the classical linear regression model) is relaxed and replaced by a very general distribution family. Within this new framework, the systematic part of the model is expanded to allow not only the mean (or location) but all the parameters of the conditional distribution of the outcome to be modelled as parametric functions of independent variables and/or latent variables. GAMLSSs have found application in health research as well as the study of floods and wildlife. For instance, Scala et al. [11] conducted a regression analysis on river floods in the Sicily Region using a GAMLSS, employing both stationary and non-stationary analyses by varying the covariates and comparing the results. They emphasized that natural events could not be adequately explained by stationary distributions alone, as they fail to account for the natural variability in variables such as temperature, atmospheric pressure, precipitation, and floods that are affected by climate change. This highlighted the utility of GAMLSS in modelling environmental data. The study also demonstrated that using physical parameters (such as rainfall) as covariates to model the distribution's parameters yielded more reliable results than solely relying on time as a covariate. Furthermore, GAMLSSs have been employed to study the impacts of fires in Portugal [12] and to investigate the variability or changes in fish species due to biological or environmental factors [13,14].

However, the effect of the independent variables on the outcome could vary according to different sea conditions, driven by a latent variable.

The two outcomes considered in our analysis, meteorological tides and significant wave height, are often analysed independently in the literature. Here, instead, we link them through a common latent structure, to avoid the quite strict assumption on their independence. The introduction of the latent structure, modelled through finite mixtures, allows us to model heterogeneity in the outcomes' distributions, often overlooked by other existing analyses. In other words, we assume that the outcomes are drawn from multiple sea conditions summarized by a fixed number of clusters, and that the effects of the independent variables are cluster-specific. We extensively discuss the role and physical meaning of the clusters underlying the observed data, obtained as a by-product of the finite mixture approach. Ignoring the underlying clustering structure would lead to a misleading inference. Regression modelling is often considered to link the two outcomes to independent environmental variables; we extend the simple linear Gaussian regression approach to further investigate outliers and, at the same time, we show the importance

of modelling not only the expected value of the outcomes but also their variances, as the independent variables also play a crucial role in affecting the variability in the data.

By clustering sea conditions, we can visualize how the outcomes behave throughout the year and identify distinct patterns. Thus, we consider a finite mixture (of regression) approach to define a unified umbrella under which multiple data features can be jointly modelled. Additionally, employing finite mixture models enhances the model's flexibility, as increasing the number of mixture components enables the approximation of any data shape.

The Italian Institute for Environmental Research and Protection (ISPRA) maintains a network of buoys to monitor wave direction and height at various points of the Italian seas. A network of ISPRA tide gauges, located along the Italian coast, additionally provides data about environmental variables. Here, we analysed data from the Gulf of Ancona, in the Adriatic Sea, as they also have specific data features due to the orography of the site [15]. However, the Mediterranean Sea, in general, is widely recognized as one of the most vulnerable regions to climate change, primarily due to its exposure to climate events [16]. In fact, temperatures in the Mediterranean Sea have been observed to be 1.4 °C higher than the global warming trend [17]. Thus, our proposal can be further applied to other sites as it contributes to shed light on the understanding of the behaviour of meteorological tides under heterogeneous conditions and bad points, which are very likely to arise in empirical data analysis but are often overlooked by most of the employed models.

Numerous models have been proposed to analyse marine physical parameters, with particular emphasis on waves and tidal currents as they represent significant hazards to coastlines. Across them, the Simulating WAVes Nearshore (SWAN) model has been widely used for analysing currents and wave behaviour. Introduced by Booij et al. [18] as a third-generation wave model, SWAN describes waves using the density spectrum and computes the evolution of the action density. Over the years, this model has been extensively employed, such as by Guillou [19], who utilized SWAN as the foundation for studying the effects of tidal currents on waves by coupling their interactions. Marsooli and Lin [20] also employed a SWAN model coupled with another model called ADCIRC to simulate storm tides and waves caused by tropical cyclones. These models are computationally complex and require substantial quantities of historical data (usually a decade [21]). Historical tidal data have been modelled to understand sea level trends. For instance, Foster and Brown [22] analysed sea level time series using simple ordinary least squares regression to estimate parameters and applied autoregression models to account for data auto-correlation, compute parameter variances, and estimate the number of degrees of freedom. Morucci et al. [23] examined wave distributions using the generalized Pareto distribution, modelling various parameters including mean, variance, and shape, to analyse extreme events along the Italian coast related to wave height.

While previous studies have analysed waves and meteorological tides separately using rather standard statistical models, our research focuses on coupling the effects of two outcomes using a bivariate model. In practice, we assume that the two outcomes are conditionally independent given the latent variable, i.e., the clusters. To accomplish this, following [15], we employ a finite mixture of GAMLSSs, defining a bivariate regression model whose parameters are cluster-specific, and this helps to identify sea regimes, allowing the identification of distinct shapes that data assume under latent environmental conditions. By employing mixture models, clusters are identified as mixtures of distributions, providing insights into the locations and shapes of clusters. Other examples are given in [24–27]; see [28] for a comprehensive review on mixture models. Parameter estimates are obtained by maximum likelihood, employing an expectation–maximization algorithm.

The paper is structured as follows: the first section introduces GAMLSSs and finite mixtures models, then a description of the coastline of Ancona is provided along with an explanation of the marine data, highlighting some features characterizing the coast taken into consideration; then, empirical analyses and results are presented; and finally, the last section presents the conclusion and suggests future developments of the study.

2. Methodology

Let $\mathbf{Y}_i = (Y_{i1}, Y_{i2}; t = i, \dots, I)$ denote sequences of bivariate observations, where Y_{i1} and Y_{i2} correspond to the meteorological tides and significant wave height a time t . Moreover, let C_i be a discrete latent variable, i.e., a cluster variable, defined on the state space $(1, \dots, k, \dots, K)$. We also have sequences of independent variables \mathbf{X}_i , including atmospheric pressure, wind speed, and fetch, that would be used to explain the sequences of the outcomes \mathbf{Y}_i . Here, in each cluster k , we are interested in modelling the conditional distributions $f(Y_{i1} | \mathbf{X}_i, C_i = k; \theta_{ik1}) = f_{ik1}$ and $f(Y_{i2} | \mathbf{X}_i, C_i = k; \theta_{ik2}) = f_{ik2}$. Three different probability density distributions as candidate distributions for the data at hand were considered: Gaussian (i.e., ignoring the bad points), log-normal, and t -family distribution. The empirical analysis was run within the R software (R-4.3.2) environment.

The resulting model was then a finite mixture of regressions [2,28]:

$$f(\mathbf{Y}_i) = \sum_{k=1}^K \pi_k \prod_{j=1}^2 f(Y_{ij} | \mathbf{X}_i, C_i = k; \theta_{ikj}) = \sum_{k=1}^K \pi_k \prod_{j=1}^2 f_{ikj}.$$

As any other model, some assumptions were made:

- Observations are independent, that is, a particular value in the data set is statistically independent of the rest. This means that its value is not influenced by any other observations.
- The population is composed of a finite number of unobserved sub-populations (K).
- We assume an endogenous clustering in an unsupervised setting, i.e., it is not possible to observe a priori to which cluster an observation belongs.

The assumption of independence among observations is a topic of ongoing debate, with some studies adopting it while others do not. In fields like wave and tide generation modelling, widely used approaches such as the Simulating WAve Nearshore (SWAN) model (Booij et al., 1999) [18] and WAve Model (WAM) (WAMDI Group, 1988) [29] are deterministic, meaning they rely on predefined functions and formulas. Consequently, these deterministic models consistently produce the same output given identical input parameters, as they lack random variables or probabilistic elements. This predictability arises from well-defined rules or equations governing the system's behaviour. In contrast, our approach employs finite mixture models under probabilistic assumptions, acknowledging the inherent uncertainty in parameter estimates. This probabilistic framework enables a more nuanced understanding of system dynamics by considering the variability in the data and providing measures of uncertainty associated with model estimates. Additionally, the assumption of independence among variables is commonly used in statistics due to its simplicity. This approach has been used for example by Lagona and Picone [30] and [15] in clustering different regimes of wave height (see also [15,24,31,32] for further examples).

The considered conditional distributions had a different number of parameters. Under the Gaussian and log-normal distributions, $\theta_{ikj} = (\mu_{ikj}, \sigma_{ikj})$, whilst the t distribution had three parameters, and $\theta_{ikj} = (\mu_{ikj}, \sigma_{ikj}, \nu_{ikj})$. In the regression framework defined under the GAMLSS setting, we linked the parameters to the independent variables as

$$\mu_{ikj} = \mathbf{X}_i' \beta_{kj}$$

$$\log(\sigma_{ikj}) = \mathbf{X}_i' \delta_{kj}$$

whilst no regression was considered for the ν_{ikj} parameters. As it was easy to depict, we assumed the same independent variables in the $J = 2$ equations and for the distributions' parameters; of course, this could be relaxed with minor efforts.

To maximize the likelihood function $\ell = \prod_i f(\mathbf{Y}_i)$, the expectation–maximization (EM; [33]) algorithm was used.

3. Marine Data Description and Modelling

The research focused on the Gulf of Ancona, located on the east coast of Italy in the Marche region (Figure 1). The port of Ancona is situated on the Adriatic Sea, which is a semi-closed sea in the northern part of the central basin of the Mediterranean Sea. This particular region of the Mediterranean Sea has garnered significant interest in numerous studies due to its size, being the largest arm of the Mediterranean Sea, and its connection to six countries. In the northern part of the Adriatic Sea, a large number of rivers and their deltas contribute sediments that are transported along the coast by currents, adding to the dynamic nature of this sea [34].

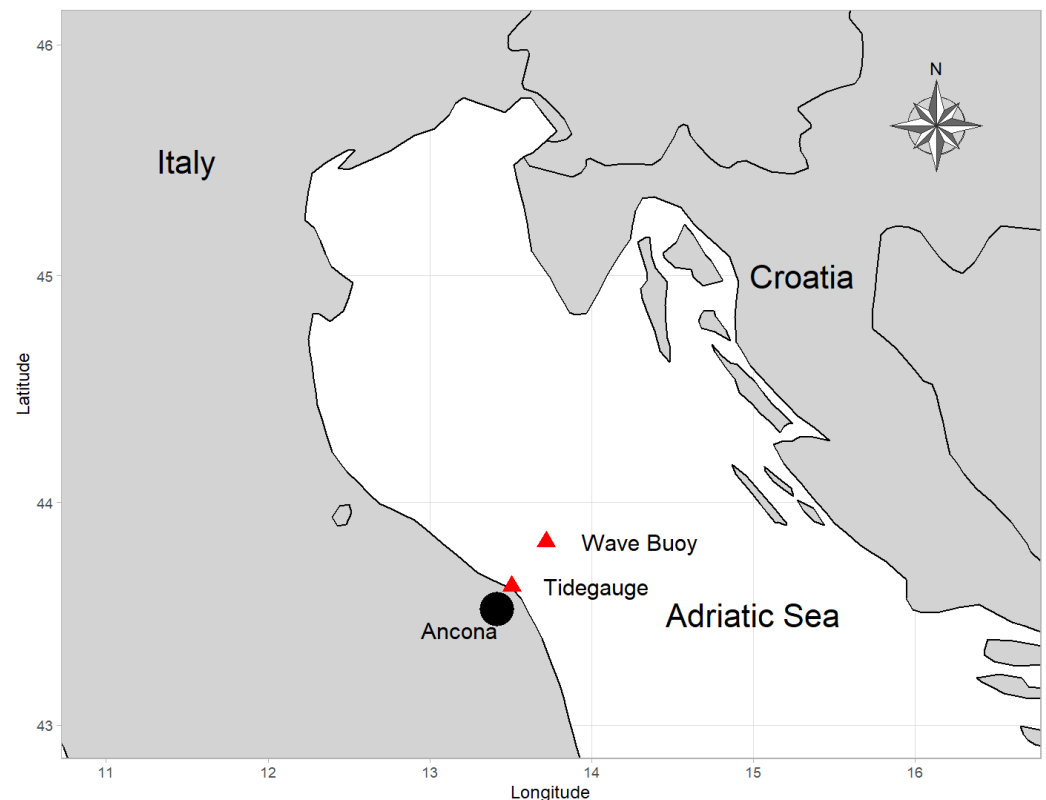


Figure 1. Italian Adriatic coast highlighting the tide gauge and the buoy of Ancona.

The prevailing winds along the Adriatic coast are predominantly from the southeast and northeast, including the scirocco, mistral, bora, and tramontane winds. The exposure to these strong winds, coupled with the sea's depth variation from around 1200 m in the south to less than 30 m in the north, leads to mixed currents and unpredictable behaviour [34]. For example, the Gulf of Venice exhibits a depth ranging from 100 m to 200 m south of Ancona [35]. Furthermore, some literature suggests that tidal currents in the Adriatic Sea are influenced by Ionian tides, as the two regions of the Mediterranean Sea converge in the Gulf of Otranto, south of Italy. Along the western side of the Adriatic Sea, the Western Adriatic Current (WAC) flows along the coastline from north to south. However, the northern part of the Adriatic Sea, including the shores of Ancona, experiences complex and variable tides. The currents exhibit a permanent counter-clockwise rotation, possibly reflecting a continuation of the behaviour of Ionian currents and tides [36].

As expected, the Adriatic Sea is also affected by sea level rise due to the impact of climate change. However, in this case, the consequences primarily manifest in the alteration of the coastal landscape. As mentioned earlier, the western coast of the Adriatic Sea, particularly the Italian shoreline, is characterized by numerous rivers and their deltas, with the Po River being the most significant. Sea level rise leads to modifications in the morphology of these landscapes and the incursion of saltwater into inland areas.

This contamination adversely affects ecosystems, soils, agriculture, and fisheries, causing significant damage [37].

In particular, the Gulf of Ancona exhibits an elbow-shaped configuration, exposed to the north, while being shielded by Monte Cornero to the south and the Apennines Mountains to the west. The prevailing wave regime in this area is predominantly influenced by winds from the northeast and southeast directions [38].

The data used for the Ancona site were obtained from both the tide gauge and the buoy, provided by the National Tide Gauge Network and the National Data Buoy Network, respectively.

Ancona is among the few locations in Italy where both a tide gauge and a buoy are installed. The tide gauge station, situated in the seaport (latitude $43^{\circ}37'29.16''$ N, longitude $13^{\circ}30'23.46''$ E), measures atmospheric pressure, air and water temperature, and relative humidity every hour. It also records wind speed and wind direction every 10 min, as well as the water level.

On the other hand, the wave buoy (latitude $43^{\circ}49'26''$ N, longitude $13^{\circ}43'10''$ E) is located offshore and collects data at 30 min intervals. It provides information on significant wave height, average wind direction, average wind speed, and other parameters that were not included in the analysis.

The availability and reliability of these data are ensured by the Italian Institute for Environmental Protection and Research (ISPRA), which coordinates and manages these monitoring networks.

Using these data, we could determine that during the analysed period (February 2021–February 2022), Ancona experienced a maximum significant wave height of 4.18 m. The significant wave height is defined as the average height, from trough to crest, of the highest one-third of the waves. It has been extensively studied in environmental research, particularly in the context of harnessing energy from the sea [39,40].

Tides are also of interest, as their behaviour is significantly influenced by strong winds and high waves. In the period under analysis, Ancona recorded a maximum meteorological residual of 114.78 cm. The meteorological residual represents the deviation between the actual tide level and the effect of the astronomical component on the sea level. In other words, it reflects the impact of meteorological parameters on the sea level. This residual is primarily influenced by atmospheric pressure, wind speed and direction, and wave height. Generally, Ancona exhibits higher tidal peaks during winter compared to summer, as depicted in Figure 2.

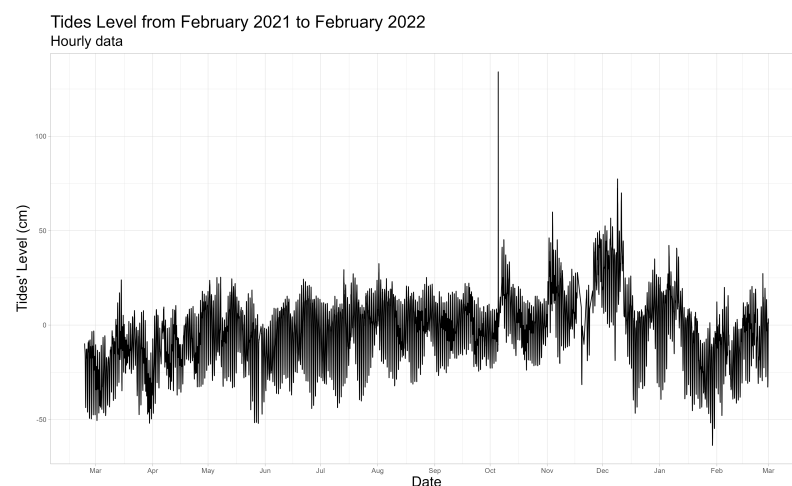


Figure 2. Hourly representation of tide levels from February 2021 to February 2022 collected by the Ancona tide gauge.

To see how much effect the meteorological component has on the sea level, here are some graphs representing both the tide level and the meteorological residual. In general, we can see that there is always a difference between the two measures.

As it is possible to observe in Figure 3, during winter, the difference between the two components is small; this is because the weather mainly contributes to the formation of tides. On the other hand, during summer the difference is more evident; this is because the formation of tides is caused mainly by other factors (astronomical components) while the weather is usually milder.

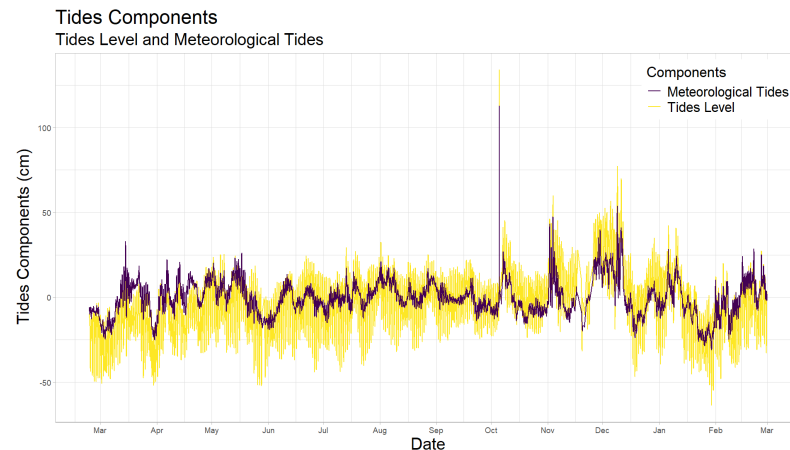


Figure 3. Plot representing the meteorological tides and the level of tides registered.

Data are collected using sophisticated technologies. In Italy, there is a network of detection tools spread all around the country. There are two different systems for collecting marine data, these are the *Rete Mareografica Nazionale* or *National Tidegauge Network (RMN)* and the *Rete Ondametrica Nazionale* or the *Italian Data Buoy Network (RON)*. The former is well spread all around Italy with 36 measurement stations which collect data every 10 min, while the latter has 15 buoys that measure and analyse data every 30 min [41]. To carry out the research, data were taken into consideration starting from 23 February 2021 to 28 February 2022. Moreover, to simplify the calculation, data were aggregated by hour, forming a data set of 8658 observations.

The analysis was conducted using the R software. The data set included several variables of interest, namely, the meteorological tides, significant wave height, atmospheric pressure, wind speed and direction, and fetch. Hourly aggregation using the mean was applied to the data, except for wind direction values, which were circular data representing the angle of the wind direction. To calculate the mean of wind direction values, the circular mean function was utilized, which computes the mean based on the cosine and sine of the angle.

Additionally, each wind direction value was associated with its corresponding fetch. The fetch was divided into 72 categories, with each category representing a 5° angle class. It is important to note that wind direction data were converted from degrees to radians, as direction was represented by an angle.

To calculate the meteorological tides, the TideHarmonics package was employed [42]. This package enables the harmonic analysis of tidal and sea level data, estimating over 400 harmonic tidal components. In the plot below (see Figure 4), it is possible to have a clear view of the three different components.

The data visualization includes three lines: the yellow line represents the values collected by the station, the purple line represents the estimated astronomical values, and the green line represents the meteorological tides, which are the components of atmospheric pressure and wind. It is evident from the observations that the values of the meteorological tides are relatively high, indicating a significant difference between the observed values and the astronomical component. This implies that the meteorological component had a substantial impact on the observed tide. Based on this observation, it can be inferred that during the initial period of the analysed year, the sea state was rough, likely due to consistently high wind speeds and/or low atmospheric pressure. Furthermore, the fetch distances were scaled by a factor of 10,000 since they were originally measured in kilometres.

It is important to note that the fetch considered in the analysis was the “effective fetch”, which refers to the effective distance of open water. This choice may result in smoother outcomes. Moving forward, we use the terms “pres” and “WS” to refer to atmospheric pressure and wind speed, respectively.

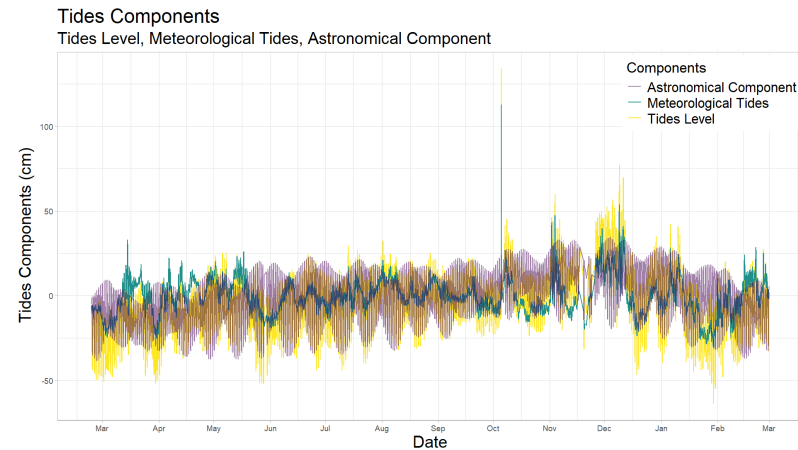


Figure 4. This plot displays the two distinct components of tides: the astronomical component, influenced by gravitational forces from the Moon and the Sun, drives the tide levels, and the meteorological tides representing the difference between tides’ levels and the astronomical component.

Two different empirical models were fitted to estimate both the mean and the variance of the two outcomes. These two empirical specifications accounted for the effective fetch in a different way. We remark again that the same specification was used for both outcomes. Firstly, we considered

$$\mu_{ikj} = \beta_{0kj} + \beta_{1kj}pres_i + \beta_{2kj}WS_i + \beta_{3kj}fetch_i, j = 1, 2; \quad (1)$$

$$\log(\sigma_{ikj}) = \delta_{0kj} + \delta_{1kj}pres_i + \delta_{2kj}WS_i + \delta_{3kj}fetch_i, j = 1, 2. \quad (2)$$

The second empirical model was:

$$\mu_{ikj} = \beta_{0kj} + \beta_{1kj}pres_i + \beta_{2kj}weights_i, j = 1, 2; \quad (3)$$

$$\log(\sigma_{ikj}) = \delta_{0kj} + \delta_{1kj}pres_i + \delta_{2kj}weights_i, j = 1, 2. \quad (4)$$

A new variable, weights, was assessed to fit the model. It took into account the wind speed and the fetch associated with the wind direction and it “weighed” them to the maximum fetch that the location under investigation had.

From now on we refer to the models as: **Model 1** and **Model 2** (Table 1). Models with a different number of clusters ($K = 2, \dots, 5$) were considered, for a total of 24 different fitted models. To compare them, three model selection criteria were used:

1. Akaike Information Criterion (AIC);
2. Bayesian Information Criterion (BIC);
3. Integrated Complete Likelihood (ICL).

Table 1. Models’ definition.

Model	Definition
Model 1	$\mu_{ikj} = \beta_{0kj} + \beta_{1kj}pres_i + \beta_{2kj}WS_i + \beta_{3kj}fetch_i$ $\log(\sigma_{ikj}) = \delta_{0kj} + \delta_{1kj}pres_i + \delta_{2kj}WS_i + \delta_{3kj}fetch_i$
Model 2	$\mu_{ikj} = \beta_{0kj} + \beta_{1kj}pres_i + \beta_{2kj}weights_i$ $\log(\sigma_{ikj}) = \delta_{0kj} + \delta_{1kj}pres_i + \delta_{2kj}weights_i$

The first criterion was assessed as:

$$AIC = -2 \times \log \ell + 2K \quad (5)$$

In this case, the log-likelihood function ($\log \ell$) is penalized by a term that depends on the number of parameters (K). The BIC follows the same pattern as the AIC but it also penalizes the log-likelihood function for the dimension of the sample size n , trying to solve the over-fitting problem.

$$BIC = -2 \times \log \ell + K \times \log n \quad (6)$$

On the other hand, the ICL criterion is assessed as the BIC penalized by the estimated mean entropy.

$$ICL = BIC - \text{entropy} \quad (7)$$

4. Results

Results for all three, Gaussian, log-normal, and t distributions, were analysed in terms of model fitting and interpretation of the estimated regression coefficients and inferred clustering. All models provided qualitatively similar results, confirming the reliability and robustness of the finite mixture approach. In terms of model selection criteria, the Gaussian mixture of regression models was preferred, confirming the idea that mixture models can accommodate different data shapes. From the AIC and BIC, the best model was the one with three clusters for Model 1 and with four clusters for Model 2, while the ICL suggested more parsimonious models, with a smaller number of clusters. Nevertheless, the differences among the criteria were not very relevant. For this reason, based on the scatter plot of the posterior probabilities we chose to look into three-cluster results for the first model and into four-component results for the second model.

As shown in Figure 5, in Model 1, the three clusters were well separated. The significant wave height and the wind speed clusters were well highlighted, and they seemed to have a different cluster-specific positive association; thus, it is clear that waves depended very much on the wind speed, but its effect was not the same over all the observations.

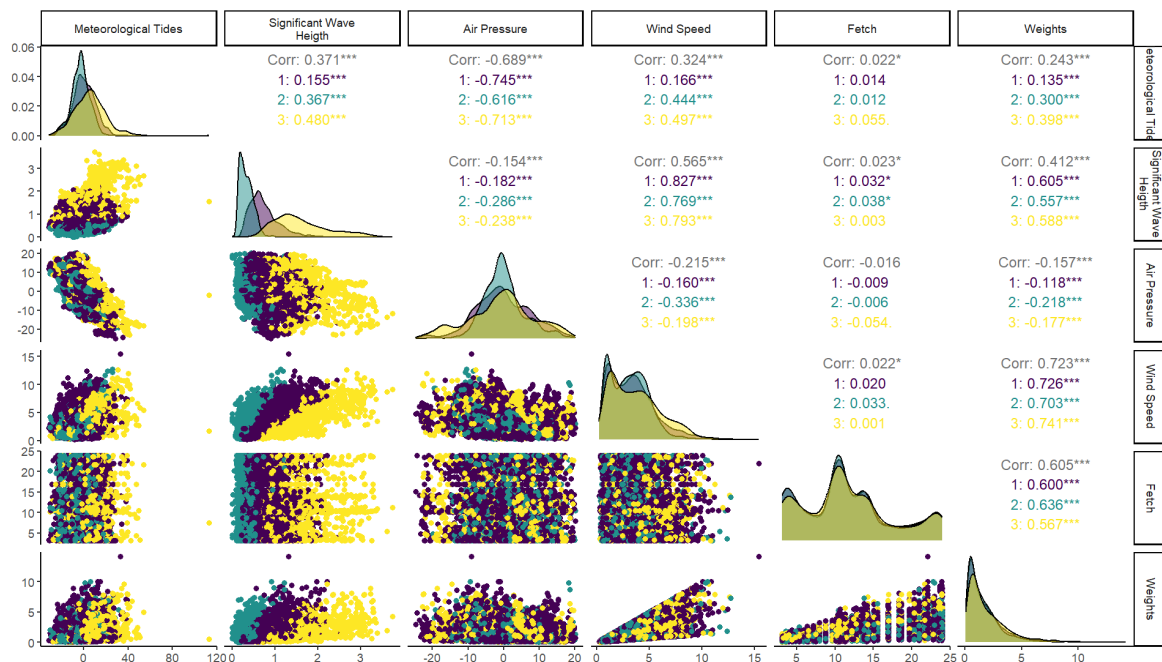


Figure 5. Clusters of the Gaussian distribution in Model 1 with three clusters. Marginal and conditional distributions for each variable are displayed on the main diagonal. The conditional correlation for each cluster is shown with the level of significance. Stars ('*' and '***') denote the level of significance.

The first cluster (purple) represented medium waves, while the second one (green) captured low waves, and the third one represented (yellow) high waves. The significant wave height was better modelled than the meteorological tides as clusters for the latter in some ways overlapped and were difficult to identify.

Let us focus on the estimated regression coefficients for both the expected value and the variance of the two outcomes.

A detailed examination of Table 2 facilitates a more nuanced interpretation of the results. Regarding the meteorological tides, it is worth noting that the estimated mean of the fetch held no significance in the model, suggesting no influence on the cluster means. This result was somehow expected as the meteorological tides are likely to be affected by meteorological factors only. We can see that the effect of atmospheric pressure was quite constant across clusters, though a reduction in its impact was estimated in Cluster 2 compared to the other two clusters. Conversely, the impact of wind speed was clearly cluster-specific, with an increasing impact on meteorological tides passing from Cluster 1 to Cluster 3.

Table 2. Estimated values of the mean and variance for the meteorological tides outcome.

	Estimated Coefficients for the Mean (μ)			Estimated Coefficients for the Variance (σ)		
	Estimate	Std. Error	<i>p</i> -Value	Estimate	Std. Error	<i>p</i> -Value
Cluster 1						
Intercept	−1.38	0.27	0.00	1.70	0.03	0.00
Atmospheric pressure	−0.98	0.01	0.00	0.01	0.00	0.00
Wind speed	0.27	0.05	0.00	0.05	0.01	0.00
Fetch	0.01	0.02	0.57	0.00	0.00	0.78
Cluster 2						
Intercept	−6.20	0.31	0.00	1.74	0.04	0.00
Atmospheric pressure	−0.80	0.02	0.00	0.01	0.00	0.00
Wind speed	1.16	0.06	0.00	0.02	0.01	0.00
Fetch	−0.01	0.02	0.79	0.00	0.00	0.58
Cluster 3						
Intercept	−4.18	0.62	0.00	2.46	0.05	0.00
Atmospheric pressure	−0.92	0.03	0.00	−0.02	0.00	0.00
Wind speed	2.23	0.09	0.00	−0.06	0.01	0.00
Fetch	0.04	0.03	0.22	−0.01	0.00	0.00

The first cluster (purple cloud on the graph) showed an intermediate situation for the meteorological residual. The atmospheric pressure had a negative effect of −0.98, which mainly drove meteorological tides. Looking at the meteorological tides–atmospheric pressure plot, we can see that within the first cluster, a strong negative relation between the two variables is evident: the more the atmospheric pressure decreases, the more the meteorological tides increase. This variable also had an impact on the variability in meteorological tides: it was statistically significant with a positive coefficient, that is, the increase in the atmospheric pressure values caused an increase in the variability in meteorological tides. Concerning the wind variable, within this first cluster, its effect was the smallest across clusters, with a significant value of 0.27. Mainly, the meteorological tides were not strongly influenced by the wind speed, but the latter had more influence on the creation of waves and currents, which indirectly had impacts on the tides. As for the atmospheric pressure, the wind speed had a positive effect on the variability in meteorological tides. Again, this may represent a situation of weather instability where the wind-speed changes influence the level of meteorological tides.

The second cluster showed a calm sea state where low values of meteorological tides were recorded. Atmospheric pressure always had a negative significant impact on the level of the meteorological tides, though smaller than the effects estimated in all other clusters, and a similar impact on the cluster specific variance as in Cluster 1. The impact of the wind

speed increased: the effect was positive with a value of 1.16. The positive effect of the wind speed in the calm sea state (Cluster 2) could be due to a lag effect, that is, there could be a time delay or lag between changes in wind speed, and their impact on the meteorological tides could not be captured by the analysis. This could lead to a situation where the effect of wind speed is positive but not yet fully realized in the meteorological tides.

The third cluster represented the rough sea, where the meteorological tides and the waves were high. The effect of the atmospheric pressure was still negative and high (-0.92), as for the previous clusters, but in this case, the effect of the wind speed was significant with a very high value of the estimated mean (2.23). As shown in the graph values, the atmospheric pressure tended to decrease rapidly and at the same time, the wind speed increased with a huge effect on the meteorological tides, hence the level of the meteorological tides was very high due to the combination of the effects of these two components. This cluster was quite interesting, as a significant effect of the fetch was estimated in the variance equation. Moreover, all the estimated coefficients in the variance equation were negative: this indicated that rough sea conditions were quite stable and persistent as long as high values of the independent variables were observed. In a nutshell, we can say that weather conditions characterizing this situation were stable and optimal to create a rough sea and the persistent or consistent state of these variables contributed to the intensity of rough seas and higher tidal levels.

Table 3 shows the estimated effects on the mean and the variance on the second outcome, the significant wave height. Bearing in mind the physical meaning of the three clusters, the impact of independent variables on the outcome was clearly cluster-specific, with the wind speed, and not the atmospheric pressure, the main determinants of the significant wave height.

Table 3. Estimated values of the mean and variance for the significant wave height response variable.

	Estimated Coefficients for the Mean (μ)			Estimated Coefficients for the Variance (σ)		
	Estimate	Std. Error	p-Value	Estimate	Std. Error	p-Value
Cluster 1						
Intercept	0.26	0.01	0.00	-2.12	0.03	0.00
Atmospheric pressure	-0.00	0.00	0.00	0.00	0.00	0.21
Wind speed	0.12	0.00	0.00	0.16	0.01	0.00
Fetch	0.00	0.00	0.02	0.00	0.00	0.65
Cluster 2						
Intercept	0.14	0.00	0.00	-3.01	0.04	0.00
Atmospheric pressure	-0.00	0.00	0.98	0.02	0.00	0.00
Wind speed	0.07	0.00	0.00	0.26	0.01	0.00
Fetch	-0.00	0.00	0.38	-0.00	0.00	0.53
Cluster 3						
Intercept	0.65	0.03	0.00	-0.95	0.05	0.00
Atmospheric pressure	-0.00	0.00	0.00	-0.01	0.00	0.00
Wind speed	0.22	0.01	0.00	0.04	0.01	0.00
Fetch	-0.00	0.00	0.76	0.00	0.00	0.71

The first cluster (purple) represented medium heights, in line with what was discussed before. The estimated mean values for this cluster were all statistically significant. The primary factor driving the formation of these medium waves was wind speed, which had a positive effect (0.12). This means that as the wind speed increased, the height of the waves also tended to increase. On the other hand, the atmospheric pressure had a negative effect on wave formation, as expected, although its influence was not particularly strong. In other words, when atmospheric pressure was high, it tended to suppress wave formation to some extent. Another significant factor was the fetch, which represents the distance over which the wind blows across the water's surface. Despite not having a high impact, it still

contributed positively to the formation of medium waves. A larger fetch generally resulted in higher wave heights. The fetch's positive influence on medium wave formation can be attributed to its capacity to provide the wind with more time and space to impart energy onto the water. This extended interaction period allows for the creation of larger waves, underscoring the importance of fetch in wave dynamics. However, in this specific case, although the estimated mean parameter for fetch appeared significant, its influence seemed to be rather small. This could be attributed to the relatively small size of the sea area under consideration; the Adriatic Sea, being an enclosed basin, has limited fetch distances.

What was particularly interesting was the estimation of the variance, as wind speed only showed a significant impact on wave variability. Thus, an increase in the wave speed led to a more variable wave height. All this was supported by the graphical inspection of the results, which showed a clear positive correlation between wave height and wind speed. In summary, the formation of medium waves in the purple cluster was primarily influenced by wind speed. Higher wind speeds contributed to increased wave height, while atmospheric pressure had a dampening effect. The fetch, although not a major factor, still contributed positively to wave formation.

Recall that the second cluster (green) represented low waves, indicating calm sea conditions. In that cluster, the estimated mean value of the wind speed was the only component that influenced the sea state. It had a slight positive effect (0.07) on the significant wave height. Although the influence of the mean wind speed estimate was not substantial, it still contributed to the formation of waves in that cluster. However, in the variance equation, both atmospheric pressure and wind speed contributed to the increase in the variability in wave heights.

The third cluster (yellow) represented high waves, indicating rough sea conditions. As already remarked referring to the meteorological tides, this cluster is particularly intriguing as it captures specific weather and sea conditions. With respect to the wave height outcome, both the atmospheric pressure and wind speed were found to be statistically significant with opposite effects. The atmospheric pressure showed a negative effect on wave height: the higher the atmospheric pressure, the lower the wave height. On the other hand, the wind speed's estimated coefficients in that cluster were the highest among the three clusters (0.22). This indicated a consistent and significant positive impact of wind speed on wave formation. In other words, as the wind speed increased, the height of the waves also increased. The presence of high waves and rough sea conditions in this cluster could be attributed to the higher wind speeds observed. Additionally, wind speed contributed to the increase in the variability in wave height, and instead, atmospheric pressure led to a lower variability. These findings suggest that the specific weather conditions characterized by a higher wind speed strongly contributes to the definition of sea conditions.

We would like to provide more details on the underlying clustering structure with a focus on seasonal patterns.

The pattern exhibited by each cluster throughout the year is depicted in Figure 6. As previously explained, the first cluster being the middle one with a general average behaviour, there was no specific seasonal pattern that characterized this kind of sea state, which was confirmed to be the steady state. The second cluster represented low meteorological tides and waves; indeed, the atmospheric pressure and the wind speed were not so high. Observations for this kind of sea state were concentrated in the spring–summer period. The third cluster highlighted the highest values for both the outcomes. These values were associated with the lower values of the atmospheric pressure and the higher values of wind speed, well describing the behaviour of the sea under specific weather conditions. What is more interesting is the period captured for this kind of phenomenon, the autumn–winter season, when we expect the sea to be rough due to adverse weather conditions.

As said before, the second analysis considered four components. Recall that in that analysis, the wind speed and the fetch were incorporated into a new index: weights.

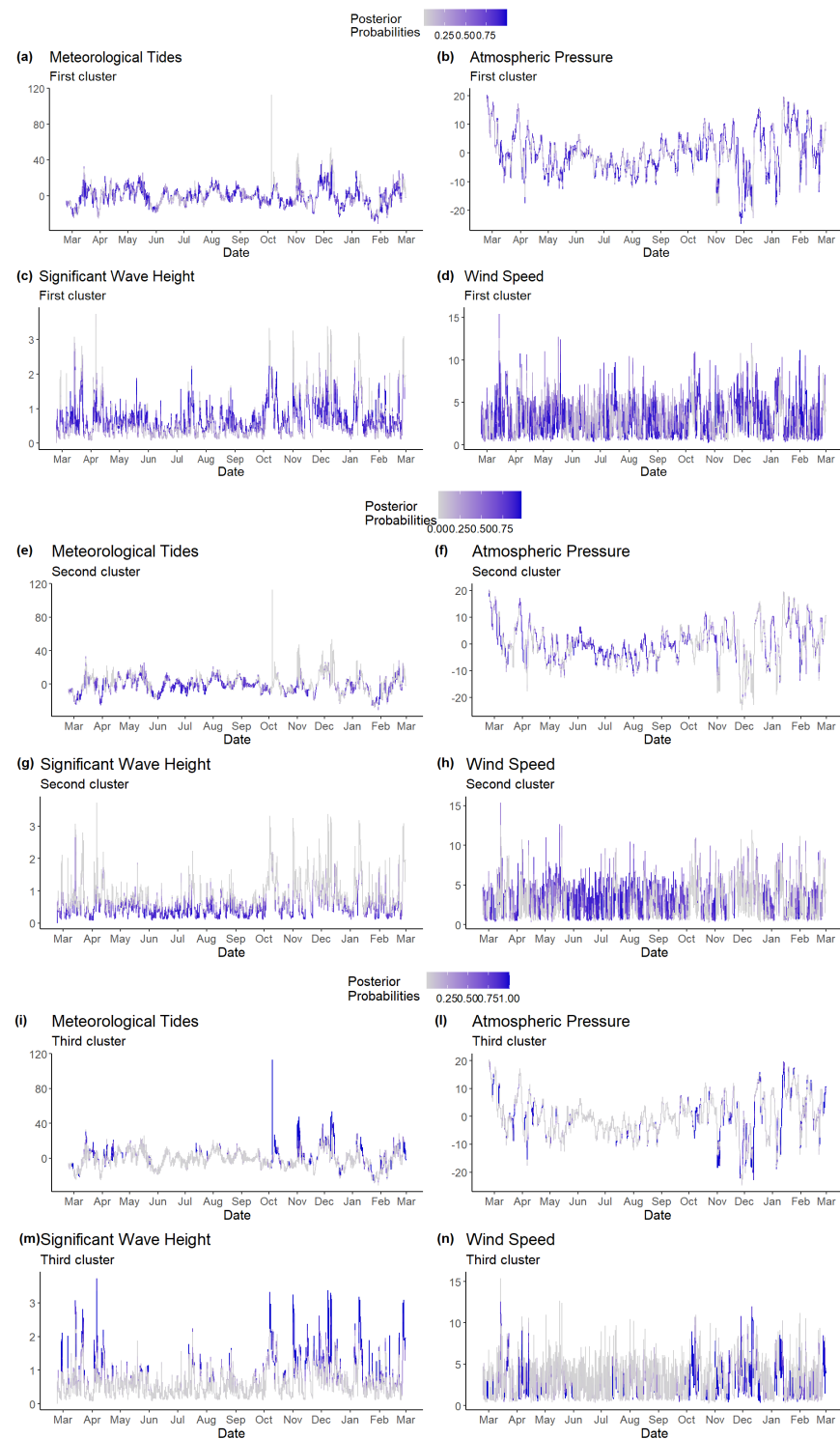


Figure 6. Clustering over time for three clusters.

Looking at Figure 7, the four clusters can be easily identified for the significant wave height. Still, the residuals' clusters are not well separated. In this case, the interpretation of the sea states identified is not straightforward, even if we can state that the second (blue) represents low waves, and the third (green) high waves, the first (purple) and fourth (yellow) components represent medium waves but in different weather conditions.

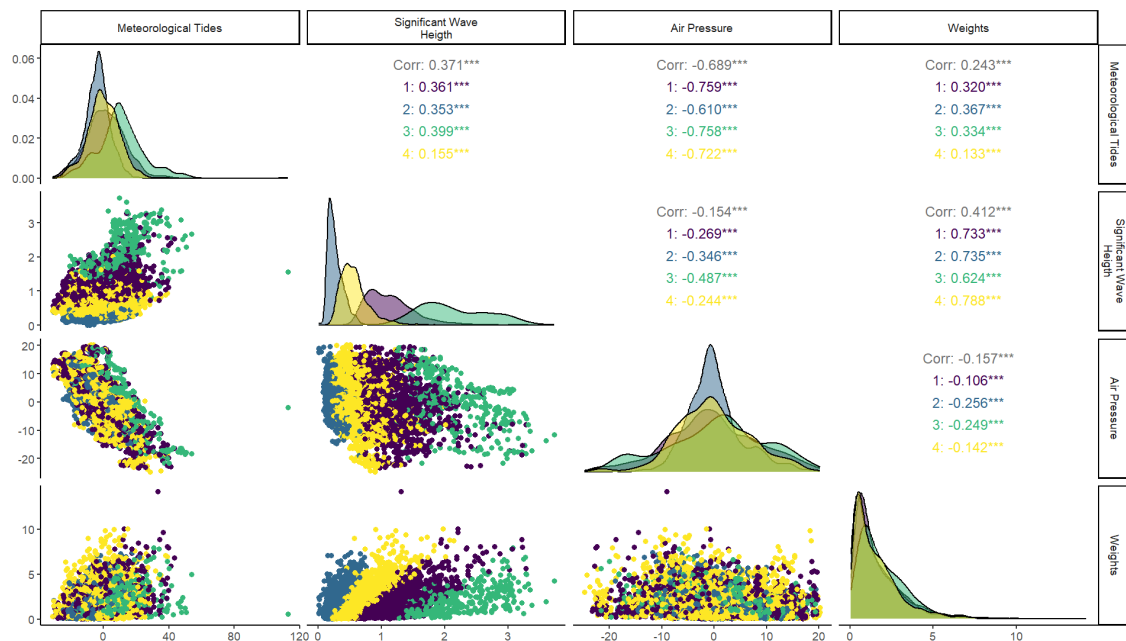


Figure 7. Clusters of the Gaussian distribution in Model 2 with four components. The diagonal of the matrix shows the distribution of each cluster for each variable. On the right side, the correlation coefficient for each cluster is shown with its level of significance. Stars ('***') denote the level of significance.

In Table 4, the mean values estimated for the meteorological tides are the same for all four clusters: both the atmospheric pressure and the weights are significant with negative and positive impacts, respectively. Although, the atmospheric pressure always has the same effect on the response variable while the weights variable seems to drive the classification.

Table 4. Estimated values of the mean and variance for the meteorological tides response variable.

	Estimated Coefficients for the Mean (μ)			Estimated Coefficients for the Variance (σ)		
	Estimate	Std. Error	p-Value	Estimate	Std. Error	p-Value
Cluster 1						
Intercept	−2.30	0.23	0.00	1.95	0.02	0.00
Atmospheric pressure	−1.00	0.02	0.00	0.00	0.00	0.65
Weights	1.83	0.11	0.00	0.01	0.01	0.42
Cluster 2						
Intercept	−5.54	0.20	0.00	1.74	0.02	0.00
Atmospheric pressure	−0.82	0.03	0.00	0.01	0.00	0.00
Weights	1.24	0.10	0.00	0.03	0.01	0.00
Cluster 3						
Intercept	4.82	0.73	0.00	2.56	0.05	0.00
Atmospheric pressure	−0.92	0.05	0.00	−0.02	0.00	0.00
Weights	2.45	0.21	0.00	−0.22	0.02	0.00
Cluster 4						
Intercept	−0.93	0.16	0.00	1.79	0.02	0.00
Atmospheric pressure	−0.97	0.02	0.00	0.00	0.00	0.01
Weights	0.23	0.08	0.01	0.05	0.01	0.00

The first cluster (purple cloud on the graph) represented a range of medium to high residuals. Within that cluster, the atmospheric pressure had a significant negative impact (−1.00) on the sea level. This is evident from the graph, which shows a pronounced decrease in atmospheric pressure leading to an elevation of the meteorological tides. In other words,

a lower atmospheric pressure was associated with higher sea levels, resulting in larger residuals. Simultaneously, the weights variable had a substantial positive effect (1.83) on the response variable. The weights variable contributed significantly to the increase in sea level. It represented the combined influence of various factors, such as wind speed and fetch, on the response variable. Higher values of the weights variable corresponded to larger sea levels, leading to increased residuals. Furthermore, the estimated parameters of the variance were not significant within that cluster. This indicated that the variability in the residuals was not influenced by the two covariates.

The second cluster (blue) represented a situation where the sea is calm, resulting in low meteorological tides. Within that cluster, the atmospheric pressure had a significant negative impact on the residuals (-0.82). The low values of the meteorological tides indicated that a high pressure was exerted on the sea surface. In other words, when the atmospheric pressure is high, it tends to suppress wave activity, leading to calmer sea conditions and smaller residuals. On the other hand, the weights variable had a high positive effect (1.24) within that cluster. This suggested that the weights played a significant role in influencing the meteorological tides. Also, the estimated parameter of the variance was significant even if not particularly high (0.03), meaning that the weights variable had a slightly positive impact on the meteorological tides' variability.

The third cluster (green) represented a situation of storms or surges, characterized by the highest values of meteorological tides. Within that cluster, the atmospheric pressure consistently had a significant negative estimated value (-0.92), indicating its strong impact on the meteorological tides. Although the values of the atmospheric pressure varied greatly, ranging from negative to positive, the negative estimated value suggested that it played a crucial role in the state of the sea and the resulting high values of the meteorological tides. The high waves associated with storms or surges likely contributed to that state. Additionally, the component of the weights in that cluster had a large and significant estimated value (2.45). This indicated that it had a substantial influence on the meteorological tides in that situation. Moreover, the estimated parameter of the variance had a negatively significant meaning where an increase in the weights values caused a decrease in the meteorological tides' variability.

The fourth and final cluster (yellow) represented a situation characterized by low-to-medium meteorological tides. Within that cluster, both estimated coefficients for the mean values were negative and positive, respectively. This suggested that the effects of these variables on the meteorological tides were not consistently in the same direction. In this scenario, the influence of the weights variable was slightly smaller (0.23). This suggested that in that particular situation, the strength of the wind was not as significant and did not have a substantial influence on the state of the sea. It implied that the wind was relatively calm and did not contribute significantly to the observed meteorological tides or wave heights. The variance of the estimated coefficients for both the atmospheric pressure and weights variables within that cluster were significant and positive. This indicated that an increase in the two covariates' values led to an increase in the meteorological tides' variability.

As for the previous analysis, the effects of the two meteorological variables were clearer when looking at the significant wave height. The latter was very well divided according to the different sea states identified.

For every cluster, the estimated coefficient of the mean was statistically significant for both explanatory variables: negative for the atmospheric pressure and positive for the weights (Table 5). Though the values estimated for the atmospheric pressure were not so high and did not differ from each other, this suggested the classification was led by changes in the weights variable.

The first cluster (purple) represented an intermediate situation characterized by medium and high waves. In this scenario, the atmospheric pressure had a relatively minor influence on the sea state. The estimated values for both the mean and variance were

close to -0.01 , indicating that changes in atmospheric pressure had a minimal impact on wave generation within this cluster.

Table 5. Estimated values of the mean and variance for the significant wave height response variable.

	Estimated Coefficients for the Mean (μ)			Estimated Coefficients for the Variance (σ)		
	Estimate	Std. Error	p-Value	Estimate	Std. Error	p-Value
Cluster 1						
Intercept	0.69	0.01	0.00	−1.31	0.02	0.00
Atmospheric pressure	−0.01	0.00	0.00	−0.01	0.00	0.00
Weights	0.20	0.01	0.00	0.11	0.01	0.00
Cluster 2						
Intercept	0.18	0.00	0.00	−2.89	0.03	0.00
Atmospheric pressure	−0.00	0.00	0.00	0.01	0.00	0.01
Weights	0.07	0.00	0.00	0.29	0.01	0.00
Cluster 3						
Intercept	1.61	0.04	0.00	−0.68	0.06	0.00
Atmospheric pressure	−0.02	0.00	0.00	−0.02	0.00	0.00
Weights	0.22	0.02	0.00	−0.07	0.03	0.01
Cluster 4						
Intercept	0.36	0.00	0.00	−2.10	0.02	0.00
Atmospheric pressure	−0.01	0.00	0.00	−0.00	0.00	0.27
Weights	0.12	0.00	0.00	0.19	0.01	0.00

On the other hand, the weights had a slightly greater positive effect for both the estimated mean and the estimated variance (0.20 and 0.11). Indeed, from the plot, we can see that there was a positive relation between the waves and the weights variable indicating that the higher the value of the weights, the higher the significant wave height. At the same time, the variability in the response variable was influenced by the weights component.

Cluster 2 (blue) represented low waves. In this scenario, the impact of atmospheric pressure was negatively significant but close to zero, indicating that atmospheric pressure had little influence on wave height. However, the variance parameter was positively significant (0.01) but small, suggesting a low impact on the variability in wave heights. The effect of wind on wave height was positive but relatively small (0.07), and the variance was also positively significant (0.29). This implies that wind contributed slightly to wave generation in that cluster, and the variability in wave height was caused mainly by wind conditions. These results, then, may represent a situation of a smooth sea, where the atmospheric pressure is low and stable, while the wind is not blowing very strong but may be changing its direction or strength, as the graph shows an increase in the weights values.

The third group (green) represented rough sea conditions with high waves. In this scenario, the atmospheric pressure had a small but significant negative impact on the significant wave height (-0.02), indicating that a higher atmospheric pressure was associated with slightly lower wave heights. The variance remained negatively significant (-0.02), suggesting that the wave height's variability was relatively low, with the model effectively capturing the relationship between atmospheric pressure and wave heights, resulting in a consistent wave behaviour. On the other hand, wind had a significant positive impact on wave height (2.45). The plot reveals that even a small increase in the weights values led to considerably higher waves, indicating that the wind in this scenario was strong and blowing in the appropriate direction to generate high waves. This suggested that the fetch, which refers to the distance over which the wind blows, was likely to be substantial. However, the variance estimate was negatively significant (-0.22), indicating that wind conditions had an impact on the wave heights' variability. Specifically, it suggested that an increase in weights values were associated with a decrease in the variability in significant wave height. This implies that the weights variable, as shown in the graph, played a role in

stabilizing the significant wave height, contributing to a more consistent pattern of rough sea conditions as long as its values remained high.

Cluster 4 (yellow) represented sea conditions characterized by low-to-medium waves. In this case, only the estimated mean of the atmospheric pressure was negatively significant (-0.01), while the estimated variance was not. The plot indicates a slight negative relationship between atmospheric pressure and wave height, suggesting that a lower pressure was associated with higher waves. On the other hand, the weights variable had both the estimated mean and variance parameters positively significant (0.12 and 0.19). The estimated mean indicated a positive effect on wave heights. Additionally, the positively significant variance suggested that the weights variable had a substantial influence on the variability in the data, indicating that fluctuations in this variable contributed to the variability observed in significant wave heights. This implies that changes in the weights variable led to fluctuations in wave heights, contributing to the overall variability in sea conditions.

Based on the previous findings, Figure 8 provides a comprehensive view of the patterns exhibited by each cluster over the course of the year. The first cluster, as previously described, represented an agitated sea state. This was evident from the significant wave height and meteorological tides, which consistently demonstrated high values within that cluster. Moreover, the atmospheric pressure consistently registered lower values, while the weights components exhibited moderate values. It is noteworthy that this agitated sea state scenario persisted throughout most of the year, with the exception of June, July, and September. This indicates that for the majority of the year, the Adriatic Sea maintained an intermediate sea state. The second cluster represented a distinct scenario characterized by a very calm sea state. In that cluster, both meteorological tides and wave heights exhibited consistently low values. Additionally, the atmospheric pressure tended to show higher values, while the weights variable remained relatively constant. This particular scenario was predominantly observed during specific months, namely, May, June, July, and August. It aligned with our expectations, as these months corresponded to the summer season when weather conditions generally promote a calm sea state. The third cluster represented a distinct set of weather conditions, as evidenced by the relatively limited number of observations captured within it. Both meteorological tides and wave heights displayed remarkably high values. Meanwhile, the atmospheric pressure exhibited considerably low values, while the weights variable did not appear to have particularly high values. This finding may seem contradictory to our theoretical understanding of sea behaviour. However, considering how the weights variable was calculated ($\frac{WS \cdot fetch}{max(fetch)}$), we can gain some insights. Low values of the weights variable may be attributed to high, near-maximum values of the fetch, representing solely wind speed values. Indeed, in this scenario, the substantial impact of the weights variable on the formation of very high waves and meteorological tides can be linked to consistent wind blowing from specific directions where fetch is significantly high. These directions likely include north or southeast winds, predominant during the highlighted period on the graph, typically occurring in the autumn–winter season. These winds encounter a significant expanse of the sea, allowing them to build waves. This interaction over expansive fetch distances contributes to the heightened impact of the weights variable on wave formation and meteorological tides within that cluster. Also, upon closer examination of the plot for the weights variable, we observed that the highlighted values often occurred immediately after or before peaks in its values were registered. This suggested a time lag in the analysis, indicating that changes in wind speed and their subsequent impact on the meteorological tides and wave heights might not have been fully captured by the current analysis. This time lag in the analysis implied that the effects of changes in wind speed on the sea state might take some time to manifest in the observed meteorological tides and wave heights. As a result, the analysis may not have accurately captured the immediate influence of wind speed on the sea conditions. The fourth and final cluster depicted a sea state characterized by low-to-medium sea levels. In that cluster, the atmospheric pressure tended to assume high values, resulting in increased pressure exerted on the sea surface. As a consequence, the sea level was lower. However,

despite the elevated atmospheric pressure, the sea state in that cluster was not calm. This can be attributed to the noteworthy variability and relatively high values of the weights factors. These factors indicated the significant influence of wind components on the sea state. Consequently, even with a high atmospheric pressure, the actions of the wind played a substantial role in maintaining a non-calm sea state. Interestingly, this particular sea state pattern was observed consistently throughout the year, suggesting its typical occurrence in the Gulf of Ancona. Regardless of the season, the fourth cluster consistently represented the prevailing sea conditions in the region.

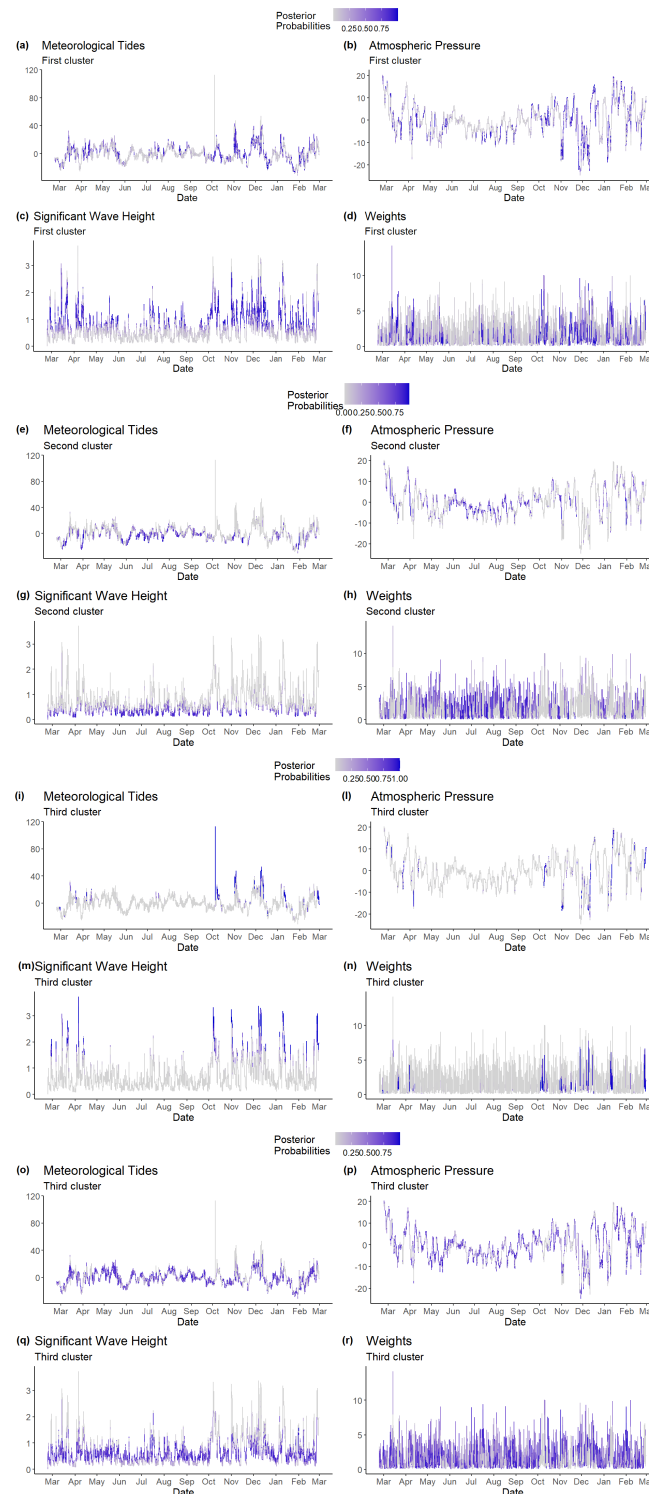


Figure 8. Clustering over time for four clusters.

In summary, both the three-component and four-component models exhibited strong fits to the data, offering valuable insights into sea state classification. The three-component model, particularly, demonstrated superior classification capabilities, as depicted in the previously described figures (Figure 5). Conversely, the four-component model, incorporating wind speed, wind direction, and fetch into a single variable (weights), yielded a four-cluster classification, representing the optimal solution. This outcome can be attributed to the intricate relationships among the meteorological components encompassed within the weights variable, which exert significant influence on sea levels. Notably, in the Adriatic Sea, tides and waves are predominantly influenced by wind dynamics. The incorporation of this composite variable enhances the model's accuracy in discerning various sea states under distinct weather conditions, with a specific emphasis on wind-related factors.

Moreover, traditional models prevalent in the literature often incur substantial computational expenses and necessitate extensive data inputs. Notably, deterministic models like the Simulating WAve Nearshore (SWAN) model rely on input data generated from other predictive models and demand high-resolution atmospheric and physical parameters due to their inclusion of spatial components. Consequently, their accuracy is contingent upon the precision of meteorological models [43]. Furthermore, the complexity of the data utilized requires more computational time. In contrast, our approach leverages raw data collected directly in situ from two distinct stations: the tide gauge and buoy located in Ancona. By adopting generalized additive models for location scale and shape (GAMLSSs), we introduced a method that enabled a flexible modelling of intricate relationships with fewer computational resources compared to conventional techniques. This capability renders GAMLSS well suited for managing extensive datasets and preventing computational overhead.

5. Conclusions

In conclusion, the research findings provided valuable insights that aligned with our theoretical understanding of sea state behaviour, particularly in the context of the Adriatic Sea. By employing the GAMLSS approach and thoroughly analysing the various meteorological components, we successfully identified and associated their effects on the sea surface. This segmentation of effects allowed for the classification of sea level rise conditions, distinguishing between normal conditions and those conducive to surges. Considering the model selection criteria, the three- and four-component models demonstrated the best performance, with slight differences observed among the results. However, the three-component model utilized in the initial analysis was selected as the most suitable choice due to its ability to effectively capture and represent the different states of the sea. This can be observed in the cluster graphs where the highlighted clusters corresponded well with the various sea conditions. Furthermore, when incorporating the variable weights, which encompassed important meteorological factors such as wind speed, wind direction, and fetch, the four-cluster classification emerged as the most appropriate. This result highlights the interconnected nature of these meteorological components and their significant influence on sea levels, particularly within the Adriatic Sea. Given the unpredictable nature of tides in this region, the wind components play a major role in shaping the sea state and influencing wave formation. The inclusion of the weights variable allowed the model to accurately identify and classify distinct sea states based on different weather conditions, with a particular emphasis on wind speed and direction. Our findings are corroborated by previous studies focusing on sea state classification. For instance, Lagona and Picone [15] employed a latent class approach to classify sea state regimes, specifically focusing on the Adriatic Sea. Their analysis revealed that the optimal classification comprised four states according to the BIC and three states according to the ICL. Additionally, their study associated the different states with distinct wind regimes and seasonal patterns. These findings align closely with our results, providing further evidence of consistency. Additionally, Bulla et al. [24] present evidence supporting the classification of sea states using hidden Markov models. Their results indicate that both three and four clusters effectively

represent the classification of sea states, delineating distinct weather conditions, with a particular emphasis on wind factors, which emerge as the primary driver in the Adriatic Sea. It is important to acknowledge that while wind speed and direction are influential factors, the applicability of these findings may vary in other geographical locations. In different regions, other factors such as the meteorological tides may also play a crucial role in determining sea levels. Consequently, the influence of atmospheric pressure becomes more significant in defining the sea state in such cases. Our analysis not only corroborate the existing literature but also provide deeper insights into the complex interplay between atmospheric conditions and sea state behaviour. By leveraging advanced modelling techniques that integrate various statistical aspects previously applied separately, we have uncovered patterns that contribute to a more profound understanding of the processes governing sea state variability. This analysis serves as a valuable foundation for future research endeavours. For instance, a further exploration could involve the inclusion of additional data from multiple years, enabling the development of a robust model for sea state forecasting. Additionally, this analysis can be expanded to investigate the effects of climate change on sea levels, offering important insights into the potential impacts of changing climatic conditions on coastal areas.

Author Contributions: Conceptualization, A.M.; methodology, A.M.; software, L.R.; validation, A.M., M.P. and L.R.; formal analysis, A.M. and L.R.; investigation, L.R.; resources, M.P.; data curation, L.R. and M.P.; writing—original draft preparation, L.R.; writing—review and editing, A.P.; visualization, L.R.; supervision, A.P.; project administration, A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been partially supported by MUR, grant number 2022XRHT8R—The SMILE project: Statistical Modelling and Inference to Live the Environment.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Aitkin, M.; Wilson, G.T. Mixture models, outliers, and the EM algorithm. *Technometrics* **1980**, *22*, 325–331. [\[CrossRef\]](#)
2. Maruotti, A.; Punzo, A. Model-based time-varying clustering of multivariate longitudinal data with covariates and outliers. *Comput. Stat. Data Anal.* **2017**, *113*, 475–496. [\[CrossRef\]](#)
3. Farcomeni, A.; Ventura, L. An overview of robust methods in medical research. *Stat. Methods Med. Res.* **2012**, *21*, 111–133. [\[CrossRef\]](#)
4. Farcomeni, A.; Greco, L. *Robust Methods for Data Reduction*; CRC Press: Boca Raton, FL, USA, 2016.
5. Davies, L.; Gather, U. The identification of multiple outliers. *J. Am. Stat. Assoc.* **1993**, *88*, 782–792. [\[CrossRef\]](#)
6. Hennig, C. Fixed point clusters for linear regression: Computation and comparison. *J. Classif.* **2002**, *19*, 249. [\[CrossRef\]](#)
7. Bai, X.; Yao, W.; Boyer, J.E. Robust fitting of mixture regression models. *Comput. Stat. Data Anal.* **2012**, *56*, 2347–2359. [\[CrossRef\]](#)
8. Rigby, R.A.; Stasinopoulos, D.M. Generalized additive models for location, scale and shape. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **2005**, *54*, 507–554. [\[CrossRef\]](#)
9. Stasinopoulos, D.M.; Rigby, R.A. Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. *J. Stat. Softw.* **2007**, *23*, 1–46. [\[CrossRef\]](#)
10. Stasinopoulos, M.D.; Rigby, R.A.; Bastiani, F.D. GAMLSS: A distributional regression approach. *Stat. Model.* **2018**, *18*, 248–273. [\[CrossRef\]](#)
11. Scala, P.; Cipolla, G.; Treppiedi, D.; Noto, L.V. The Use of GAMLSS Framework for a Non-Stationary Frequency Analysis of Annual Runoff Data over a Mediterranean Area. *Water* **2022**, *14*, 2848. [\[CrossRef\]](#)
12. Sá, A.C.; Turkman, M.A.; Pereira, J.M. Exploring fire incidence in Portugal using generalized additive models for location, scale and shape (GAMLSS). *Model. Earth Syst. Environ.* **2018**, *4*, 199–220. [\[CrossRef\]](#)
13. Colloca, F.; Enea, M.; Ragonese, S.; Di Lorenzo, M. A century of fishery data documenting the collapse of smooth-hounds (*Mustelus* spp.) in the Mediterranean Sea. *Aquat. Conserv. Mar. Freshw. Ecosyst.* **2017**, *27*, 1145–1155. [\[CrossRef\]](#)
14. Costa, E.F.; Teixeira, G.M.; Freire, F.A.; Dias, J.F.; Fransozo, A. Effects of biological and environmental factors on the variability of *Paralichthys brasiliensis* (Sciaenidae) density: An GAMLSS application. *J. Sea Res.* **2022**, *183*, 102203. [\[CrossRef\]](#)
15. Lagona, F.; Picone, M. Model-based clustering of multivariate skew data with circular components and missing values. *J. Appl. Stat.* **2012**, *39*, 927–945. [\[CrossRef\]](#)
16. Giorgi, F. Climate change hot-spots. *Geophys. Res. Lett.* **2006**, *33*. [\[CrossRef\]](#)

17. Marini, K. Climate and environmental change in the Mediterranean—main facts. *MedEC Erişim* **2018**, *1*, 2019.
18. 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](#)]
19. Guillou, N. Modelling effects of tidal currents on waves at a tidal stream energy site. *Renew. Energy* **2017**, *114*, 180–190. [[CrossRef](#)]
20. Marsooli, R.; Lin, N. Numerical modeling of historical storm tides and waves and their interactions along the US East and Gulf Coasts. *J. Geophys. Res. Ocean.* **2018**, *123*, 3844–3874. [[CrossRef](#)]
21. Sun, Q.; Little, C.M.; Barthel, A.M.; Padman, L. A clustering-based approach to ocean model–data comparison around Antarctica. *Ocean Sci.* **2021**, *17*, 131–145. [[CrossRef](#)]
22. Foster, G.; Brown, P.T. Time and tide: Analysis of sea level time series. *Clim. Dyn.* **2015**, *45*, 291–308. [[CrossRef](#)]
23. Morucci, S.; Picone, M.; Nardone, G.; Arena, G. Tides and waves in the Central Mediterranean Sea. *J. Oper. Oceanogr.* **2016**, *9*, s10–s17. [[CrossRef](#)]
24. Bulla, J.; Lagona, F.; Maruotti, A.; Picone, M. A multivariate hidden Markov model for the identification of sea regimes from incomplete skewed and circular time series. *J. Agric. Biol. Environ. Stat.* **2012**, *17*, 544–567. [[CrossRef](#)]
25. Maruotti, A.; Alaimo Di Loro, P. CO₂ emissions and growth: A bivariate bidimensional mean-variance random effects model. *Environmetrics* **2023**, *34*, e2793. [[CrossRef](#)]
26. Huang, W.; Dong, S. Probability distribution of wave periods in combined sea states with finite mixture models. *Appl. Ocean Res.* **2019**, *92*, 101938. [[CrossRef](#)]
27. Huang, W.; Dong, S. Joint distribution of individual wave heights and periods in mixed sea states using finite mixture models. *Coast. Eng.* **2020**, *161*, 103773. [[CrossRef](#)]
28. McLachlan, G.J.; Lee, S.X.; Rathnayake, S.I. Finite Mixture Models. *Annu. Rev. Stat. Appl.* **2019**, *6*, 355–378. [[CrossRef](#)]
29. The Wamdi Group. The WAM model—A third generation ocean wave prediction model. *J. Phys. Oceanogr.* **1988**, *18*, 1775–1810.
30. Lagona, F.; Picone, M. A latent-class model for clustering incomplete linear and circular data in marine studies. *J. Data Sci.* **2011**, *9*, 585–605. [[CrossRef](#)]
31. Lagona, F.; Picone, M.; Maruotti, A.; Cosoli, S. A hidden Markov approach to the analysis of space–time environmental data with linear and circular components. *Stoch. Environ. Res. Risk Assess.* **2015**, *29*, 397–409. [[CrossRef](#)]
32. Lagona, F.; Picone, M.; Maruotti, A. A hidden Markov model for the analysis of cylindrical time series. *Environmetrics* **2015**, *26*, 534–544. [[CrossRef](#)]
33. McLachlan, G.J.; Krishnan, T. *The EM Algorithm and Extensions*; John Wiley & Sons: Hoboken, NJ, USA, 2007.
34. Franco, P.; Michelato, A. Northern Adriatic Sea: Oceanography of the basin proper and of the western coastal zone. In *Marine Coastal Eutrophication*; Elsevier: Piscataway, NJ, USA, 1992; pp. 35–62.
35. Orlic, M.; Gacic, M.; Laviolette, P. The currents and circulation of the Adriatic Sea. *Oceanol. Acta* **1992**, *15*, 109–124.
36. Zonn, I.S.; Kostianoy, A.G. The Adriatic Sea. In *The Boka Kotorska Bay Environment*; Springer: Cham, Switzerland, 2017; pp. 19–41.
37. Carbognin, L.; Teatini, P.; Tosi, L. The impact of relative sea level rise on the Northern Adriatic Sea coast, Italy. *WIT Trans. Ecol. Environ.* **2009**, *127*, 137–148.
38. per la Protezione, A. *Atlante Delle Onde Nei Mari Italiani Italian Wave Atlas*; Roma Tre University: Rome, Italy, 2004.
39. Zheng, C.W.; Li, C.Y. Variation of the wave energy and significant wave height in the China Sea and adjacent waters. *Renew. Sustain. Energy Rev.* **2015**, *43*, 381–387. [[CrossRef](#)]
40. Ali, M.; Prasad, R.; Xiang, Y.; Deo, R.C. Near real-time significant wave height forecasting with hybridized multiple linear regression algorithms. *Renew. Sustain. Energy Rev.* **2020**, *132*, 110003. [[CrossRef](#)]
41. Canesso, D.; Cordella, M.; Arena, G. Manuale di Mareografia e Linee Guida per i Processi di Validazione dei Dati Mareografici. ISPRA Manuali e Linee Guida 77/2012; 2012; ISBN 978-88-448-0532-6. Available online: <https://www.isprambiente.gov.it/it/pubblicazioni/manuali-e-linee-guida/manuale-di-mareografia-e-linee-guida-per-i-processi-di-validazione-dei-dati-mareografici> (accessed on 26 April 2024).
42. Stephenson, A.G. Harmonic Analysis of Tides Using TideHarmonics. 2016. Available online: <https://CRAN.R-project.org/package=TideHarmonics> (accessed on 26 April 2024).
43. Dykes, J.D.; Wang, D.W.; Book, J.W. An evaluation of a high-resolution operational wave forecasting system in the Adriatic Sea. *J. Mar. Syst.* **2009**, *78*, S255–S271. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.