D3.6-User document for statistical analysis tools for urban-scale hazards and data samples

AUTHOR(S) – Iulia Anton (ATU Sligo), Sudha-Rani Nalakurthi (ATU Sligo), Roberta Paranunzio (CNR-ISAC), Salem Gharbia (ATU Sligo), Michele Bendoni (LaMMA), Francesca Caparrini (CNR), Alberto Ortolani (LaMMA), Carlo Brandini (LaMMA), Roberto Vallorani (LaMMA), Gianni Messeri (LaMMA)

DOCUMENT TRACKS DETAILS

Project acronym	SCORE
Project title	Smart Control of the Climate Resilience in European Coastal Cities
Starting date	01.07.2021
Duration	48 months
Call identifier	H2020-LC-CLA-2020-2
Grant Agreement No	101003534

Deliverable Information	
Deliverable number	D3.6
Work package number	WP3
Deliverable title	User document for statistical analysis tools for urban-scale hazards and data samples
Lead beneficiary	ATU Sligo
Author(s)	Iulia Anton (ATU Sligo), Sudha-Rani Nalakurthi (ATU Sligo), Roberta Paranunzio (CNR-ISAC), Salem Gharbia (ATU Sligo), Michele Bendoni (LaMMA), Francesca Caparrini (CNR-ISAC), Alberto Ortolani (LaMMA), Carlo Brandini (LaMMA), Roberto Vallorani (LaMMA), Gianni Messeri (LaMMA)
Due date	30/06/2023
Actual submission date	30/06/2023
Type of deliverable	Report
Dissemination level	Public

VERSION MANAGEMENT

Revision table			
Version	Name	Date	Description
V 0.1	Iulia Anton (ATU Sligo), Sudha-Rani Nalakurthi (ATU Sligo), Roberta Paranunzio (CNR-ISAC), Salem Gharbia (ATU Sligo), Bendoni (LaMMA), Francesca Caparrini (CNR-ISAC), Alberto Ortolani (LaMMA), Carlo Brandini (LaMMA), Roberto Vallorani (LaMMA), Gianni Messeri (LaMMA)	09/06/2023	First draft
V 0.2	Leo Creedon (ATU Sligo), Mar Riera (ENT), Jacek Barańczuk (UG), Maria Manuela Portela (IST-ID)	23/06/2023	Updated draft internally reviewed
V 0.3	Iulia Anton (ATU Sligo), Sudha-Rani Nalakurthi (ATU Sligo), Roberta Paranunzio (CNR-ISAC), Salem Gharbia (ATU Sligo), Bendoni (LaMMA), Francesca Caparrini (CNR-ISAC), Alberto Ortolani (LaMMA), Carlo Brandini (LaMMA), Roberto Vallorani (LaMMA), Gianni Messeri (LaMMA)	27/06/2023	Updated draft after contribution from partners
V1.0	Iulia Anton (ATU Sligo) Salem Gharbia (ATU Sligo)	29/06/2023	Final version

LIST OF ACRONYMS AND ABBREVIATIONS

Acronym / Abbreviation	Meaning / Full text
AD	Anderson-Darling test
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
BM	Block Maxima
CCLL	Coastal City Living Lab
CI	Confidence Intervals
DT	Digital Twin
EBA	Ecosystem Based Approach
EDF	Empirical Distribution Function
ETT	Extremal Types Theorem
EU	European Union
EVA	Extreme Value Analysis
EVT	Extreme Value Theory
EWSS	Early Warning Support System
GEV	Generalised Extreme Value
GOF	Goodness-Of-Fit
GPD	Generalised Pareto Distribution
MLE	Maximum Likelihood Estimation
MRL	Mean Residual Life plot
NBS	Nature Based Solutions
PCA	Principal Component Analysis
PDS	Partial Duration Series
POT	Peak Over Threshold
PP	Point Process approach
QQ	Quantile-Quantile plot
RCP	Representative Concentration Pathway
RL	Return Level

RPO	Research Performing Organisation
SIP	SCORE ICT Platform
SME	Small and Medium-sized Enterprises
WCSS	Within-Cluster Sum of Squares
WP	Work Package

BACKGROUND: ABOUT THE SCORE PROJECT

The intensification of extreme weather events, coastal erosion and sea-level rise are major challenges to be urgently addressed by European coastal cities. The science behind these disruptive phenomena is complex, and advancing climate resilience requires progress in data acquisition, forecasting, and understanding of the potential risks and impacts for real-scenario interventions. The Ecosystem-Based Approach (EBA) supported by smart technologies has potential to increase climate resilience of European coastal cities; however, it is not yet adequately understood and coordinated at European level.

SCORE is a four-year EU-funded project aiming to increase climate resilience in European coastal cities. The project outlines a co-creation strategy, developed via a network of 10 coastal city 'living labs' (CCLLs), to rapidly, equitably and sustainably enhance coastal city climate resilience through EBAs and sophisticated digital technologies.

The 10 CCLLs involved in the project are: Sligo and Dublin, Ireland; Barcelona/Vilanova i la Geltrú, Benidorm and Basque Country, Spain; Oeiras, Portugal; Massa, Italy; Piran, Slovenia; Gdansk, Poland; Samsun, Turkey.

SCORE will establish an integrated coastal zone management framework for strengthening EBA and smart coastal city policies, creating European leadership in coastal city climate change adaptation in line with The Paris Agreement. It will provide innovative platforms to empower stakeholders' deployment of EBAs to increase climate resilience, business opportunities and financial sustainability of coastal cities.

The SCORE interdisciplinary team consists of 28 world-leading organisations from academia, local authorities, RPOs, and SMEs encompassing a wide range of skills including environmental science and policy, climate modelling, citizen and social science, data management, coastal management and engineering, security and technological aspects of smart sensing research.

EXECUTIVE SUMMARY

This document is a deliverable of the SCORE project, funded under the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003534.

The aim of this document is to describe the processing tools and data of D3.5 (https://doi.org/10.5281/zenodo.8034107). This document includes a brief theoretical explanation of the statistical analysis tools, a description of the tools' usage, and the description of the produced datasets.

LINKS WITH OTHER PROJECT ACTIVITIES

The data produced by the tools in this task will enable other downstream models, and in particular the hydraulic and land-sea interaction models used for urban-scale flooding (Task 3.4) and also the long-term coastal evolution models (Task 3.5), whose development will build also on the main hazardous conditions at different occurrence probabilities identified by the clustering analysis. This is crucial for risk estimation in WP6 and for related Nature Based Solutions (NBS) adaptation strategies in WP7. Outputs of this Task 3.3 will finally be tested through a general analysis procedure in Task 3.6. Time series and outputs produced in this task are among the main sources of data production stored in the SCORE ICT Platform - SIP - (WP5) for the entire project and a fundamental component to drive models in Digital Twin (DT) and EWSS (Early Warning Support System) in WP8.

TABLE OF CONTENTS

INDEX OF FIGURES

INDEX OF TABLES

INTRODUCTION

The purpose of D3.6 is to provide a description of the set of tools and procedures (included in D3.5) exploited to assess from a statistical point of view the climate data retrieved by the previous Tasks 3.1 and 3.2. Building on D3.5, the aim of this document is to provide a statistically based and standardized procedure to analyse historical time series and projections derived from the previous tasks, along with step-by-step examples to apply these tools to their own case studies and data by means of a range of R functions.

These tools can be used to estimate trends in parameters of interest in coastal urban areas (sea level, waves, wind, rainfall, sea temperature, etc.), and related extreme values and to identify suitable distributions and return periods. Moreover, statistical tests, including goodness-of-fit tests are provided, along with more advanced multivariate analysis. Specifically, the main addressed points here are:

i. past, present and future trend analysis based on linear models,
ii. extreme value modelling from relevant extreme value theory,
iii. clustering techniques for the multivariate data analysis aimed at grouping homogenous conditions for coastal hazard occurrence in the CCLLs,
iv. analysis of the weather patterns over a defined region (in this case, the CCLLs areas) associated to the main coastal hazards in the CCLLs.

This activity is propaedeutic to the development of models and scenarios of hydraulic risks at different time scales, functional to a range of different interventions to address climate change adaptation and mitigation and emergency management in the CCLLs. Moreover, the statistical-based approach provided in this task is transferable to other geographic contexts to enable coastal risk scenarios identification.

Each section starts with a brief theoretical introduction on the statistical analysis approach used and followed by a description of the R tools exploited. Each section is provided with examples of application, mainly for significant wave heights ($H_s(m)$) and river discharge ($Q(\frac{m^3}{s})$).

1.TIME SERIES TREND ANALYSIS

Time series analysis is often used to estimate the trends in parameters of interest (sea level, waves, wind, rainfall, sea temperature, etc.) as exemplified in the present chapter.

Understanding the parameters of interest is very important to analyse the underlying causes of their pattern and the trend that is leading to certain consequences. With increasing climate change consequences, and available historical data, analysing the affected parameters such as sea levels, waves, wind, rainfall, etc. is crucial. Time series analysis is a simple and specific way of analysing these parameters by collecting the sequence of data over a period of time (Stasinopoulos, 2007). The collected data readings are studied through time-series analysis.

A simple additive model in the time-series can be expressed as follows (Stasinopoulos, 2007):

$$Y(t) = Trend + Seasonality + Residual.$$

Here *Y* is the parameter component that has been observed at a time-period (*t*).

For the current study, hourly significant wave height $H_s(m)$ has been observed for Marina di Massa, Italy, from the years 1979 to 2018, according to the **Table 1**.

Table 1 Total number of *Hs(m)* values from 1979 to 2018 for Marina di Massa, Italy

No.	Year	No of values	No.	Year	No of values	No.	Year	No of values
1	1979	8736	15	1993	8760	29	2007	8760
2	1980	8784	16	1994	8760	30	2008	8784
3	1981	8760	17	1995	8760	31	2009	8760
4	1982	8760	18	1996	8784	32	2010	8760
5	1983	8760	19	1997	8760	33	2011	8760
6	1984	8784	20	1998	8760	34	2012	8784
7	1985	8760	21	1999	8760	35	2013	8760
8	1986	8760	22	2000	8784	36	2014	8760
9	1987	8760	23	2001	8760	37	2015	8760
10	1988	8784	24	2002	8760	38	2016	8784
11	1989	8760	25	2003	8760	39	2017	8760
12	1990	8760	26	2004	8784	40	2018	8760
14	1992	8784	28	2006	8760			

The figure below, **Figure 1,** shows the hourly time series for $H_s(m)$ at Marina di Massa. We can observe that there is a seasonality that needs to be studied to understand the $H_s(m)$. This step suggests that further study of the parameter is required.
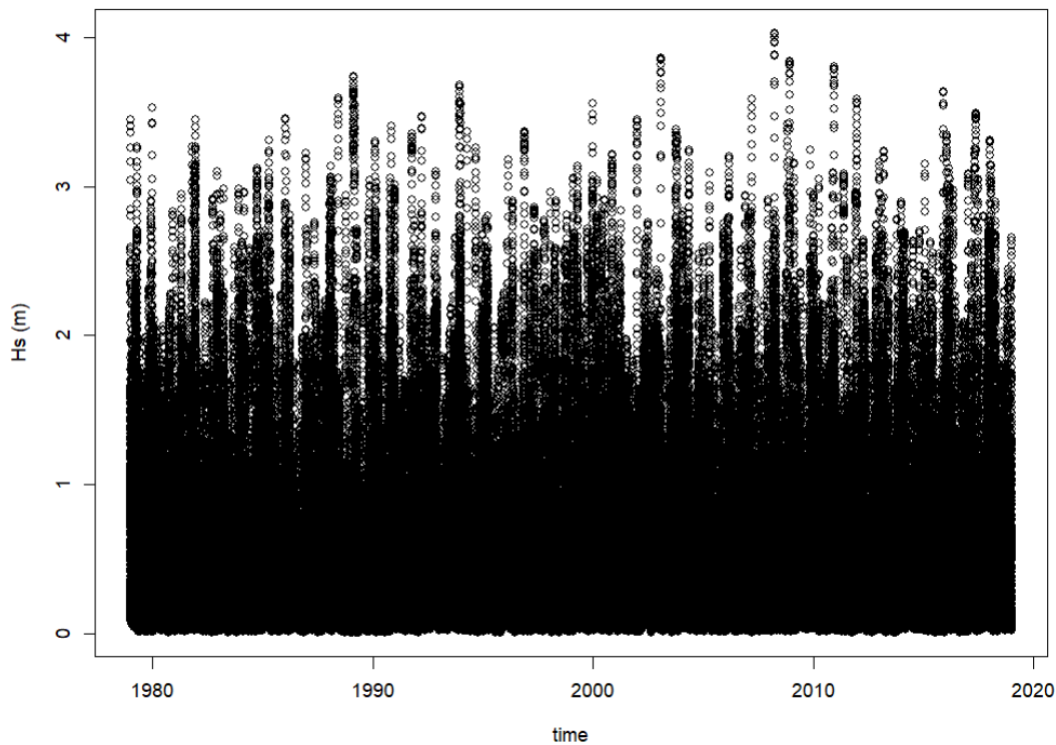
Figure 1: Time-series visualisation of the parameter Hs(m)

Figure 2 shows the mean variations for each month. This is achieved by averaging the daily data to monthly data, grouping the same month data for every year (for example, grouping all the January averages for each year to see the time-series variation).
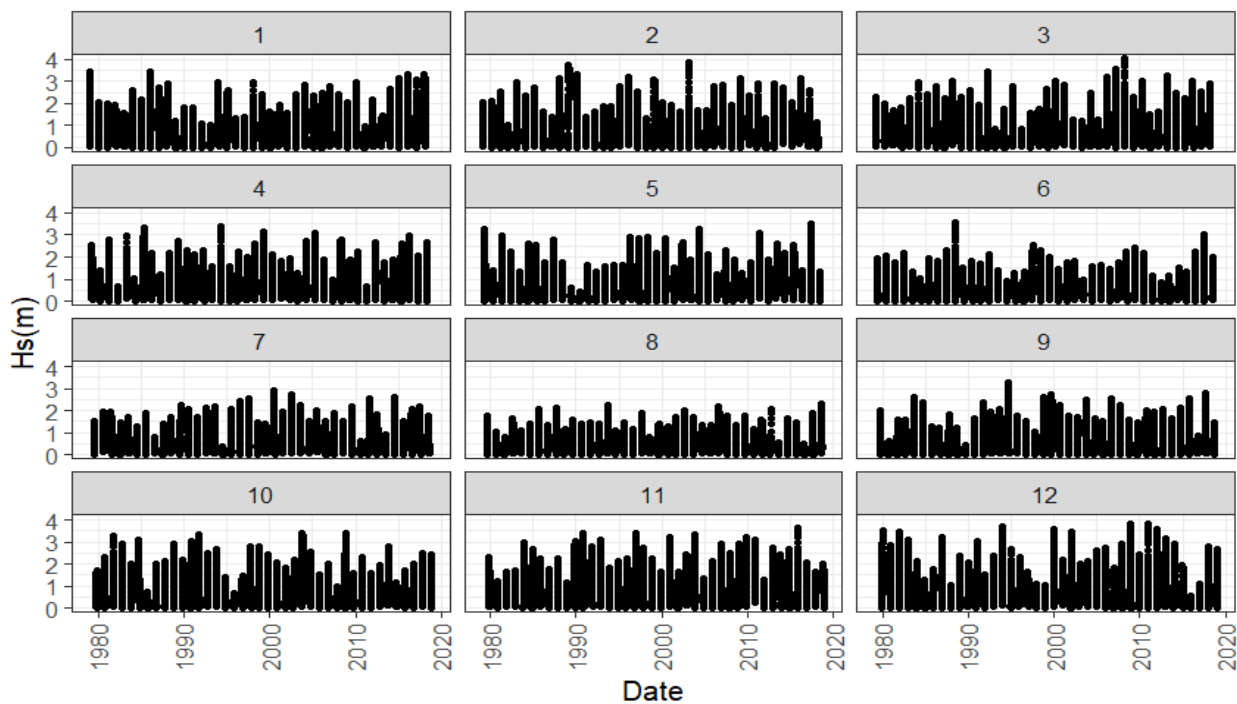
Figure 2: Monthly average patterns of the historical data; the numbers 1-12 correspond to the months of January to December

Figure 3 shows the maximum and mean values of each month for the historical time-period data, and there is an increasingtrend observed from the pattern as well (blue line).
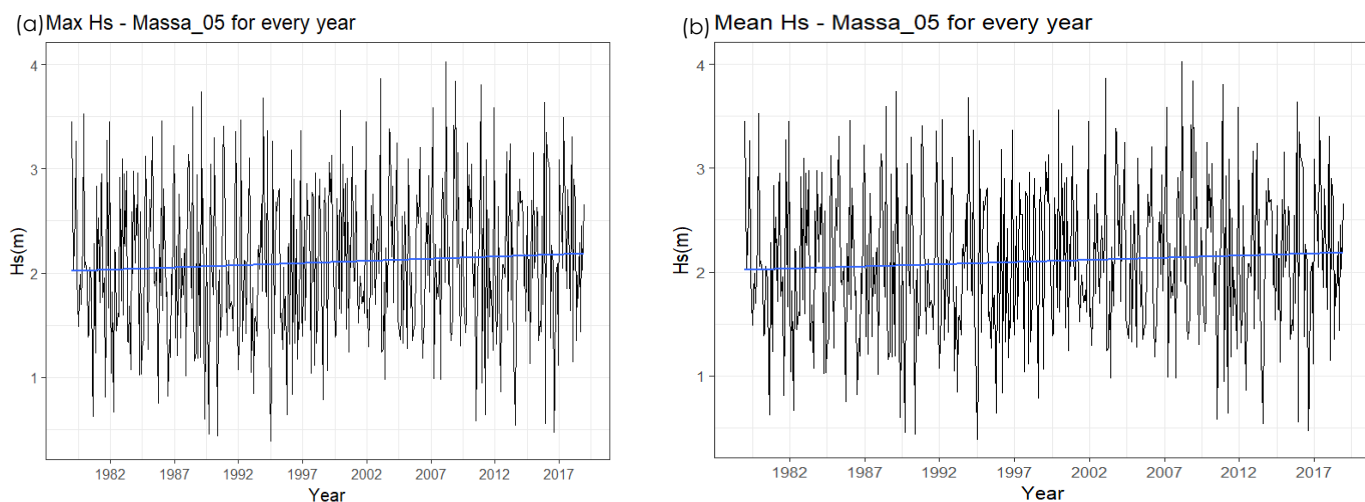
Figure 3: Monthly (a) maximum and (b) mean values of the time-series data

The figure below, Figure 4 shows the increasing trend of the $H_s(m)$ (blue line) as well for the RCP 4.5 and RCP 8.5 scenarios over the 2006-2100 time period.
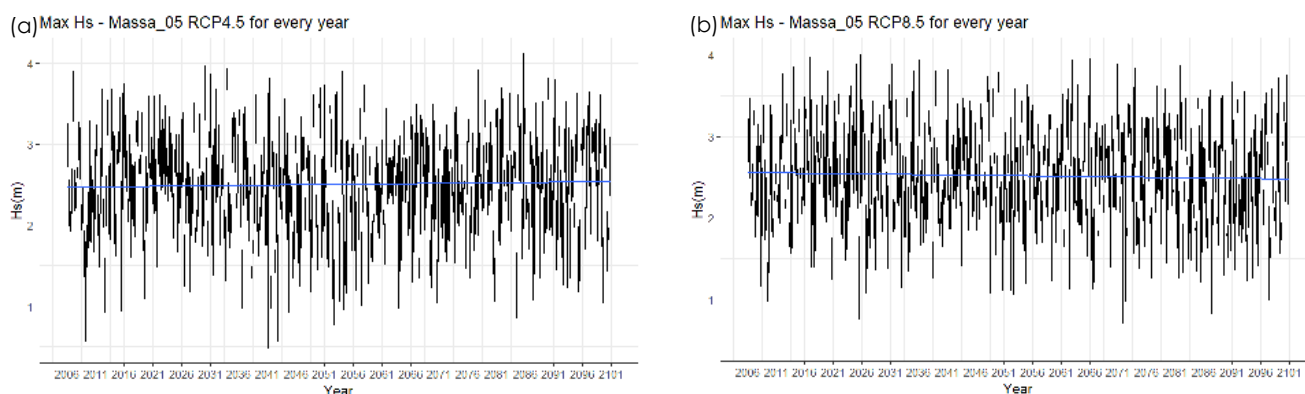
Figure 4: Monthly maximum values of the time-series data for (a) RCP4.5 and (b) RCP8.5

The significant trends, for all the data ($H_s(m)$ for historical, RCP4.5, and RCP 8.5 scenarios) were identified using the Mann-Kendall test (Kendall, 1957; Mann, 1945). The test is to detect the presence of a monotonic tendency in the time series data. For this assessment, it makes no assumptions about the underlying distribution of the data, or its rank-based measures to influence any extreme values.

The results from this test, for the hourly $H_s(m)$, shows that the test statistic is 0.029 and the corresponding two-sided $P - value$ is less than 2.22e-16. Because this $P - value$ is less than 0.05, we reject the null hypothesis of the test and conclude that a trend is present in the data, as we already saw in the before figures.

2. EXTREME VALUE ANALYSIS

2.1. Introduction and methods

The traditional way of analyzing extreme values is based on extreme values distributions. Originally introduced by Fisher & Tippett, 1928, these distributions account for maxima distributions in samples of identically distributed random and independent variables. Extreme value analysis (EVA) or Extreme Value Theory (EVT) aim to develop mathematical models and methods able to predict the occurrence of rare phenomena i.e., to estimate the likelihood of the occurrence of extreme values based on a few basic assumptions and observed/measured data (Benstock & Cegla, 2017). Although the extreme values are not frequent, the impact of these values on humans is huge. Studying such extreme values provides an overview of the parameter, for instance the rareness of such extreme data.

Typical approaches are applied to annual maxima or occasionally over a different time period (e.g., monthly) i.e., the Block Maxima (BM). The classical reference on this method is (Gumbel, 1958) though there are a range of other methods to fit the distributions. Other alternative approaches have been introduced more recently. One is to look at exceedances over a defined threshold rather than maxima over a fixed period i.e., the Peak Over Threshold (POT) method (Pickands, 1975). There is a third approach as well, which uses r-largest order statistics that considers values based on the comparison of BM and POT.

The classical mathematical formulation is based on asymptotic results analogous to the Central Limit Theorem. Let $X_1, X_2, X_3, \ldots, X_n$ be a set of random variables. Each X_i is sampled from some unknown underlying distribution function F. The aim is to study the behaviour of the maximum value (the extreme value), which can be expresses as $M_n = Max\ \{X_1, X_2, X_3 \ldots \ldots X_n\}$.

However, the distribution should divide the observation period into non-overlapping periods of equal size and only considers the maximum observations in each period (Gumbel 1958).

The general assumption for the simple maximum is that for the integer $h \geq 0$ and $n \geq 1$, the distribution of the random vector $(X_{h+1}, X_{h+2}, \ldots, X_{h+n})$ does not depend on 'h'. And for the maximum M_n, we seek the limiting distribution of $(M_n - b_n)/a_n$ for some choice of **normalising constants** $a_n > 0$ and b_n (further description is provided in (Ferro, 2005).

For estimating the **probability** (P) of an extreme value (or a rare event) above the threshold u, the general empirical frequency equation is, where N is the sample size:

$$\hat{P} = \frac{1}{N}\sum_{i=1}^{N} 1_{x_i > u}.$$

where $1_{x_i > u}$ denotes the indicator of a rare or extreme event, that exceeds the threshold value.

2.2. Block Maxima

The Block Maxima technique in extreme value theory (EVT) focuses only on the maximum observation during each non-overlapping, equal-sized interval of the observation period. The resulting new observations roughly follow an extreme value distribution under conditions of extreme value (Ferreira & De Haan, 2015). Thus, one can say that the Block Maxima (BM) or Extremal Types Theorem (ETT) addresses a specific question, which is 'for a set of j independent identically distributed positive values, $\{X_1, X_2, \ldots, X_j\}$, what are the possible limiting distributions of M_j' (Anderson et al., 1984), where $M_j = a_j\left[Max\left(X_1, X_2, X_3 \ldots \ldots X_j\right) - b_j\right]$ as $j \to \infty$.

$$F^j\left(\frac{x - b_j}{a_j}\right) \xrightarrow[j \to \infty]{} G(x),$$

where a_j and b_j are the normalising constant for a set of 'j' independent identically distributed positive values, M_j is the set of maximum values for each block, $G(x)$ is the non-degenerate distribution function, and F^j is the distribution function j' values.

There are three cumulative distribution functions known as Generalized Extreme Value (GEV) cumulative distribution functions for the Extreme Value Analysis (EVA):

1) Gumbel distribution:

$$F(x) = e^{-\left(\frac{x-\mu}{\sigma}\right)^{-\xi}} \quad x \in R,\, x > \mu,\, \xi > 0,\, \sigma > 0$$

2) Fréchet distribution:

$$F(x) = e^{-\left(\frac{x-\mu}{\sigma}\right)^{-\xi}} \quad x \in R,\, x > \mu,\, \xi > 0$$

3) Weibull distribution:

$$F(x) = 1 - e^{-\left(\frac{x}{\sigma}\right)^{\xi}} \quad x \geq 0,\, \sigma > 0,\, \xi > 0$$
$$F(x) = 0,\, x < 0$$

In these equations, μ is the location parameter, ξ is the shape parameter, and σ is the scale parameter. The shape parameter ξ governs the distribution type, also called Type I, Type II and Type III distribution functions: Type I with ξ = 0 (Gumbel distribution); Type II with ξ > 0 (Fréchet distribution); Type III with ξ < 0 (Weibull distributions). The general analysis shows that Weibull distribution fits can be found for wind, waves, and sea level parameters.

The BM guarantees that if a limit exists for maxima, it must have one of these specified forms of distribution.

The Generalized Extreme Value (GEV) distribution is the combined approach for the above three distribution functions, and it can be written as follows (cumulative distribution function):

$$F(x) = e^{-\left(1 + \xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}},$$

defined for values of x for which $\xi \cdot x > \xi \cdot \mu - \sigma$.

The Block Maxima method was tested for the $H_s(m)$ from Marina di Massa, Italy. For this analysis we used annual maximum values from the original hourly $H_s(m)$. In Figure 5, we can see the time series plot which shows that the $H_s(m)$ ranges from 2.8 m to 4 m.

The BM process is run in R-package (Gilleland & Katz, 2016), with a block size 1 year (12 months) for the model fit, and maximum likelihood estimation (MLE) is used for both GEV and Gumbel distributions. The parameters of the generalised extreme value distribution were: $\mu= 3.209$ as location, $\sigma= 0.324$ as scale and $\xi= -0.316$ as shape parameters, while standard errors for each are 0.056, 0.040, and 0.963 respectively. The shape parameter from the analysis has a negative value, therefore the fitted GEV distribution is a Weibull distribution.
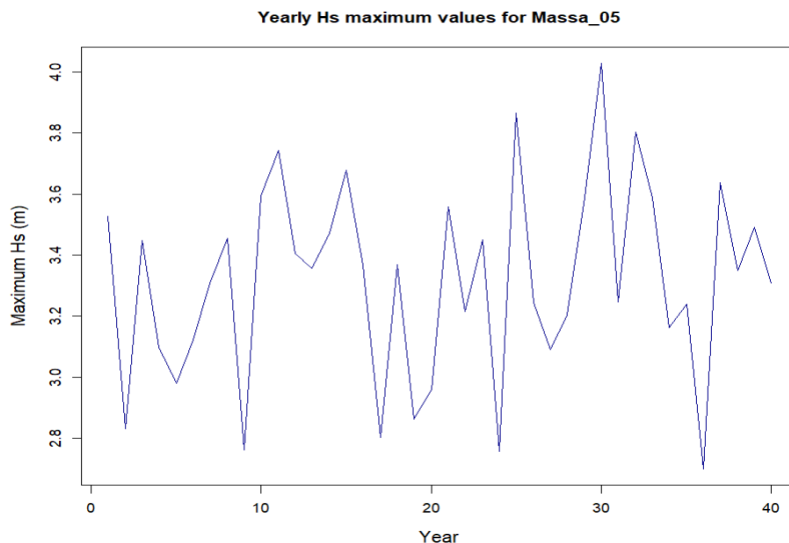
Figure 5 Annual maximum values of $H_s(m)$ for Massa, Italy

The goodness of fit was assessed graphically by various diagnostic plots shown on Figure 6. The model fitted very well to a certain level around 3.2 m to 3.8 m and it seems above that data is slightly less extreme than what the Weibull distribution would suggest, however the overall fit seems fine.
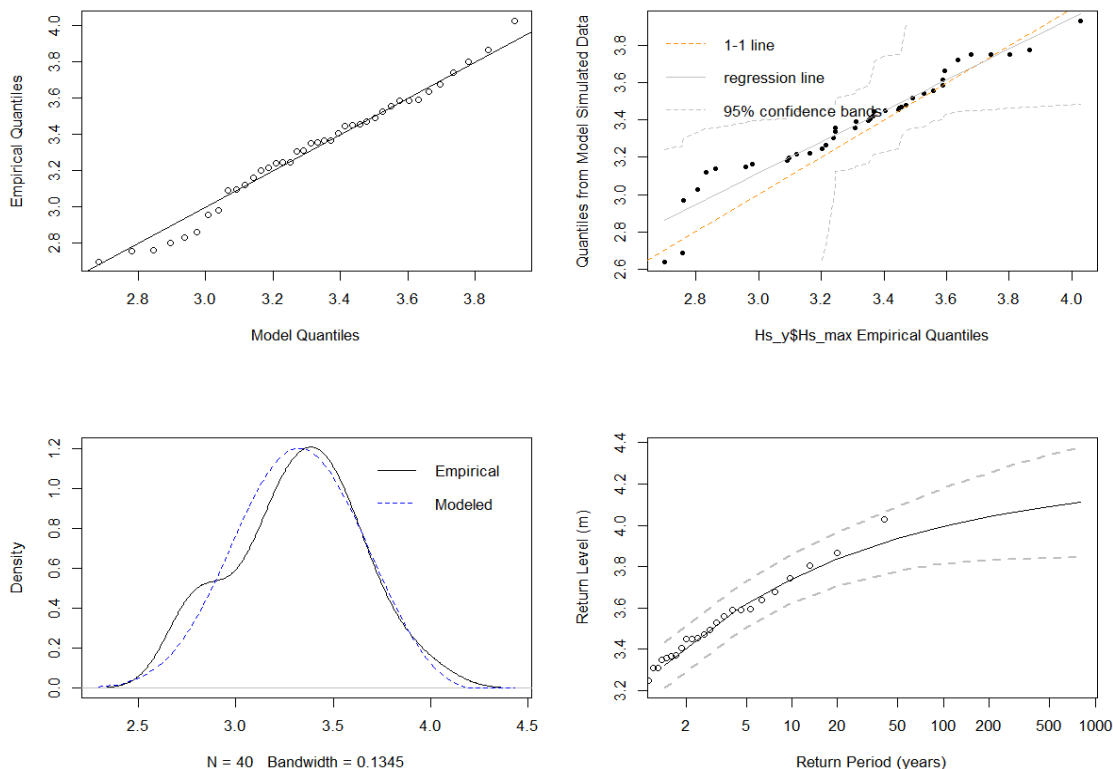
Figure 6: GEV model diagnostics plots

Figure 7 shows the likelihood and gradients in the neighbourhood of the optimum values reached for each parameter (with the other parameters held at their optimum values) (Gilleland & Katz, 2016). GEV fits seem reasonable based on the assumptions made.
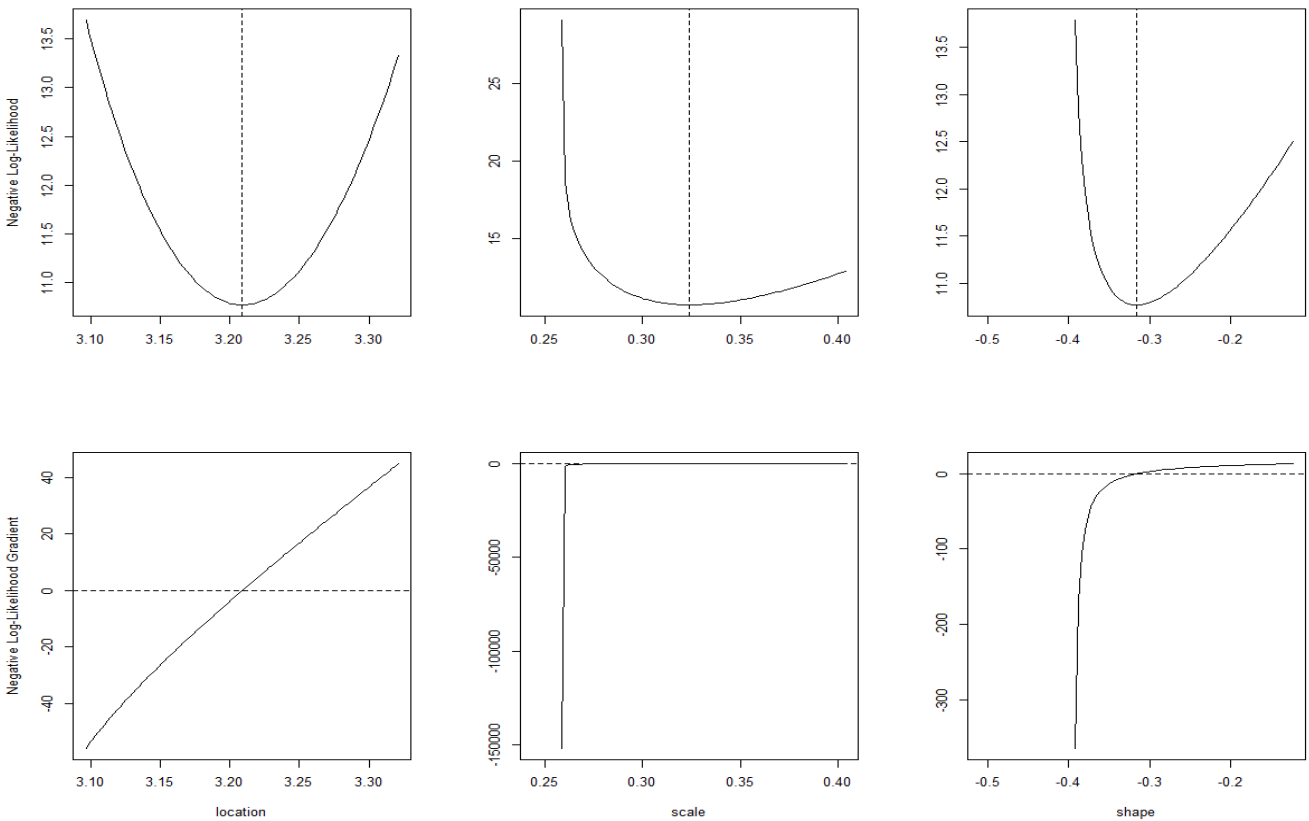
Figure 7: Location, Scale and Shape parameter likelihood gradients for GEV distribution

The Gumbel distribution has been used for the return periods and model diagnostics for the comparison with the GEV distribution. We have obtained μ= 3.156 as location, σ= 0.3086804 as scale and ξ =0 as shape. However, the model fit for the Gumbel distribution shows that the $H_s(m)$ values does not fit well with the model values (Figure 8). Figure 9 is similar to Figure 7 from the GEV distribution which shows the scale and shape factors for the Gumbel distribution, where unlike GEV the scale factor, the scale factor exceeds 0. The figures and values also suggest that the Gumbel distribution is not a good fit for the $H_s(m)$ EVT. Additionally, the best fitted GEV model can be justified, based on the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and likelihood ratio test. Lower values of AIC and BIC indicate the better model (*Table 2*) between the two distributions.

Table 2: AIC, BIC and the negative likelihood ratio tests results

	AIC	BIC	Negative likelihood ratio test
GEV	27.548	32.614	10.773
Gumbel	31.716	35.094	13.858
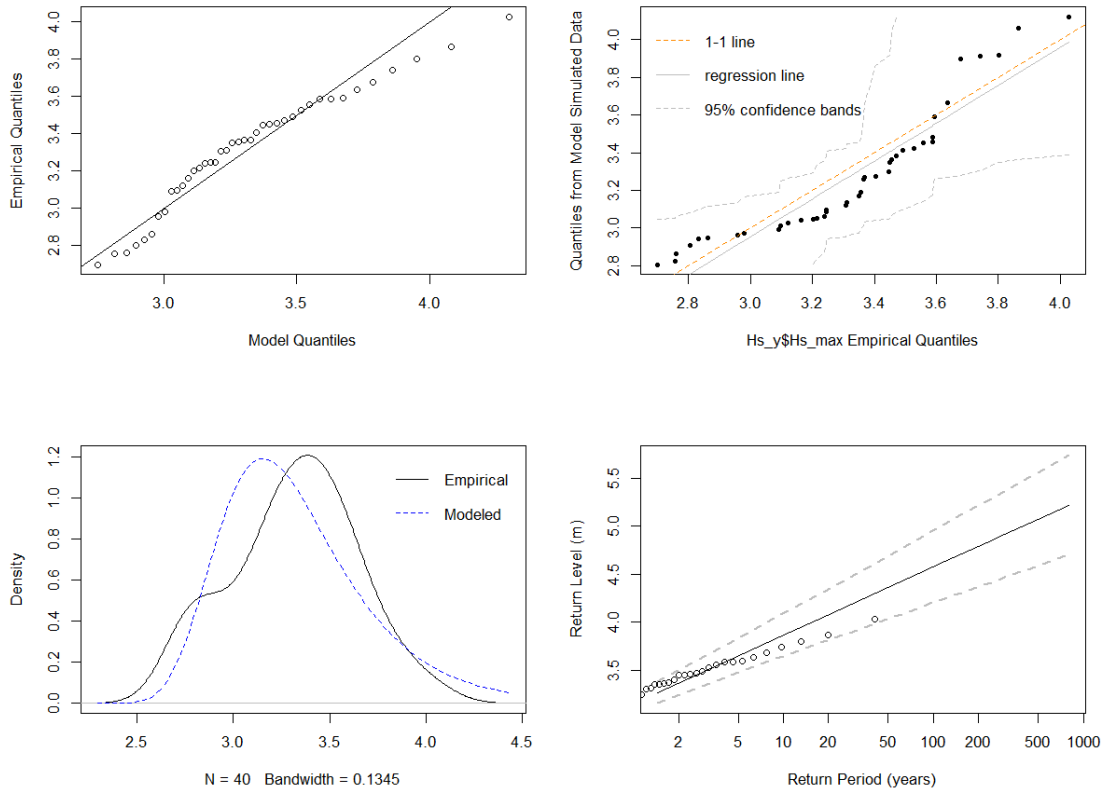
Figure 8 Gumbel distribution model fit
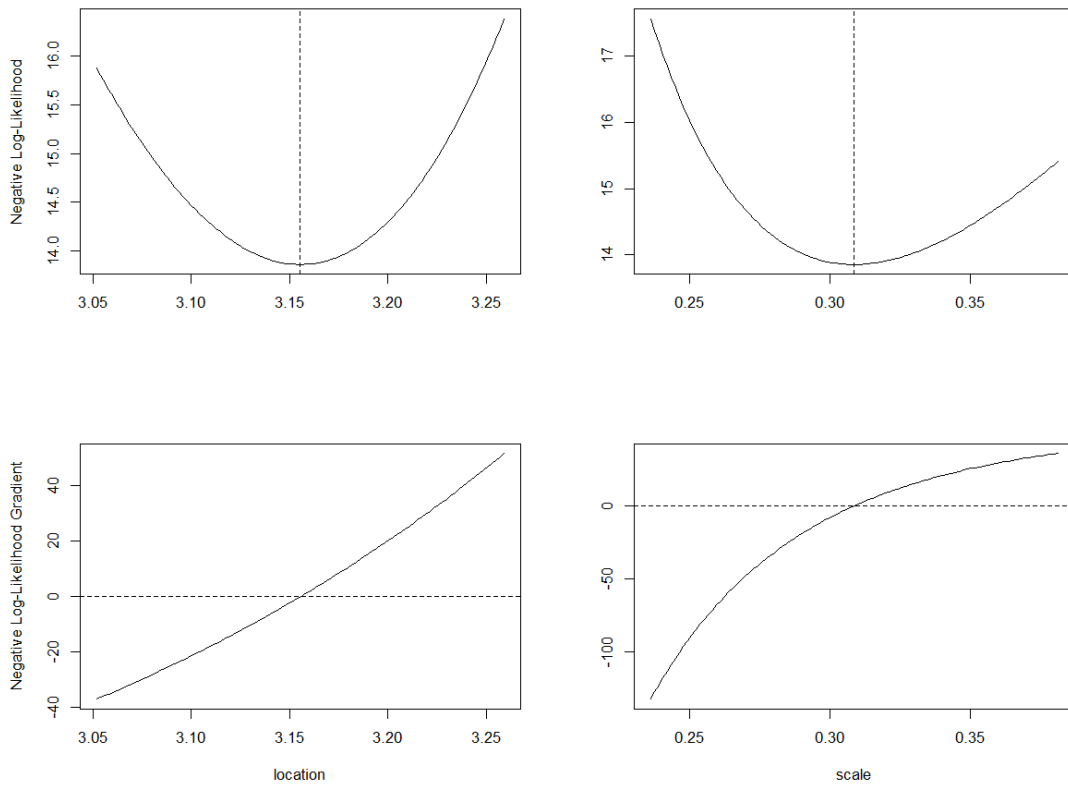
Figure 9: Location and Scale parameter likelihood gradients for Gumbel distribution (shape=0)

Results expressed in return levels for both GEV and Gumbel distribution functions are summarised in *Table 2* . These are the maximums which occur once on average during the given return period. The likelihood test also reveals, $p-value$ of 0.013 which is less than 0.05. Therefore, as expected, the Gumbel distribution is not suitable.

In **Table 3,** the return period estimates along with their Confidence Intervals calculated for both distributions are presented.

Table 3: Return levels for the $H_s(m)$

|	95% lower CI	Estimate	95% upper CI
GEV distribution			
25-year return period	3.727	3.861	3.996
50-year return period	3.779	3.936	4.092
100-year return period	3.811	3.995	4.178
200-year return period	3.830	4.042	4.255
Gumbel distribution			
25-year return period	3.864	4.143	4.421
50-year return period	4.034	4.360	4.686
100-year return period	4.202	4.576	4.949
200-year return period	4.369	4.790	5.212

2.3. Peak over Threshold

To obtain the maximum of each block, the BM approach separates the data into many blocks. To have sufficient blocks, a very large data set is required. To overcome this, a modern approach, known as the Peak over Threshold (POT) approach is elaborated in the sections below.

2.3.1.　Introduction

The POT approach works by specifying a certain high threshold and considering all observations above the threshold point in the analysis. Therefore, it is always critical to find an appropriate threshold for POT method. The POT method was first developed by (Pickands, 1975) with both theoretical framework and statistical tools (Ferreira & De Haan, 2015). Rather than block size, POT uses threshold value. Therefore, the POT method uses all the relevant high observations and neither misses any higher observations, nor retains any lower observations, thus making better use of available data. The Generalized Pareto Distribution (GPD) model is usually used for the POT.

For a sample of observations X_t for $t = 1,2, \dots, n$, a simple way to extract the extremes observations exceeding over a predetermined high threshold (u) is considered. This exceedance can be recorded when the $X_t > u$ for any $t = 1,2, \dots, n$. The cumulative distribution function for such sample observations is $F(x) = \Pr(X_t \leq x)$, the excess over u is defined as $y = X_t - u$ and this approach is called POT.

For a parameter, if the independent identically distributed random variables of a function, and the data has shown exceedance values above a specific threshold (u), the GPD function formula, in terms of cumulative distribution function (CDF) is as follows (Saeed Far & Abd. Wahab, 2016):

$$F(x) = 1 - (1 + \frac{\xi y}{\sigma_u})^{-\frac{1}{\xi}}, \qquad for\ \sigma > 0\ and\ \xi \neq 0$$

$$F(x) = 1 - \exp\left(-\frac{y}{\sigma_u}\right), \qquad for\ \sigma > 0\ and\ \xi = 0\ ,$$

where the distribution function of the excesses y is over the threshold u. Therefore, $y \geq 0$ if $\xi \geq 0$ and $0 \leq y \leq -\sigma_u/\xi$ if $\xi < 0$. Here, σ is the scale parameter, and ξ is the shape parameter.

Further description of POT and related theorems are described in (Bommier, 2014). The step-by-step procedure for the POT approach is:

1. Select the threshold u

2. Extract the exceedances from the observations

3. Fit the GDP distributions

4. Estimate the return levels.

2.3.2. Threshold selection – Graphical Method

The threshold should ensure extraction of maximum information from the time series, without violating statistical assumptions underlying the Partial Duration Series (PDS) approach (Kiran & Srinivas, 2021). PDS approach considers the events that exceed an established threshold retrieved from a time-series. The objective of this threshold is to help reduce the number of observations and reduce the variance by avoiding too low values as thresholds that induce bias (Acero et al., 2018).

There are two main ways to select a threshold value for a set of observations:

1) Graphical analysis through plots, and

2) Numerical approach via equations, which again can be sub-categorised as model fit, non-parametric models, computational approach etc.

The easiest method to estimate the residuals is to fit the Generalised Pareto Distribution (GPD) for a range of thresholds (Coles, 2001; Smith, 2002). The reparameterization of generalised Pareto scale parameter is one technique to see the estimates by avoiding the scale change $\sigma^* = \sigma_u - \xi u$. The parameter stability plot (Figure 10) shows plotting the parameter estimates ξ (shape parameter) and the reparametrized scale (σ_u) from the GPD model against the range of thresholds (u). The plots show that for threshold (u) above 1 (known as optimal threshold values), the shape and scale values are nearly constant. For the present study, the statistical modelling of extreme events from R-packages "ismev" (Janet E. Heffernan, 2022) and "extRemes" (Gilleland & Katz, et al. 2016) are used for the graphical analysis of waves.

The mean residual life (MRL) plot is another commonly used method for threshold selection (Coles, 2001). The MRL uses the mean of the GPD function for the threshold selection. The mean of the Generalised Pareto Distribution E(Y), for a random variable Y (a nonnegative random variable), with parameters σ (scale parameter), and ξ (shape parameter) is described as

$$E(Y) = \frac{\sigma}{1-\xi}\ , for\ \xi < 1.$$

For any higher threshold values v where $v > u$, the expected value is

$$E(Y - v|y > v) = [\sigma_U + \xi * (v - u)]/(1 - \xi)$$

It is linear in higher thresholds v with the intercept $[\sigma_U - \xi * u]/(1 - \xi)$ and gradient $\xi/(1 - \xi)$. The estimated mean residual life above a threshold v is given by the sample mean excess mean $(x[x > v]) - v$. The approximation process starts with a pre-chosen threshold value, then the GPD is fitted to the excesses using maximum likelihood estimation. The threshold should be chosen at the value where the shape and scale parameters remain constant.
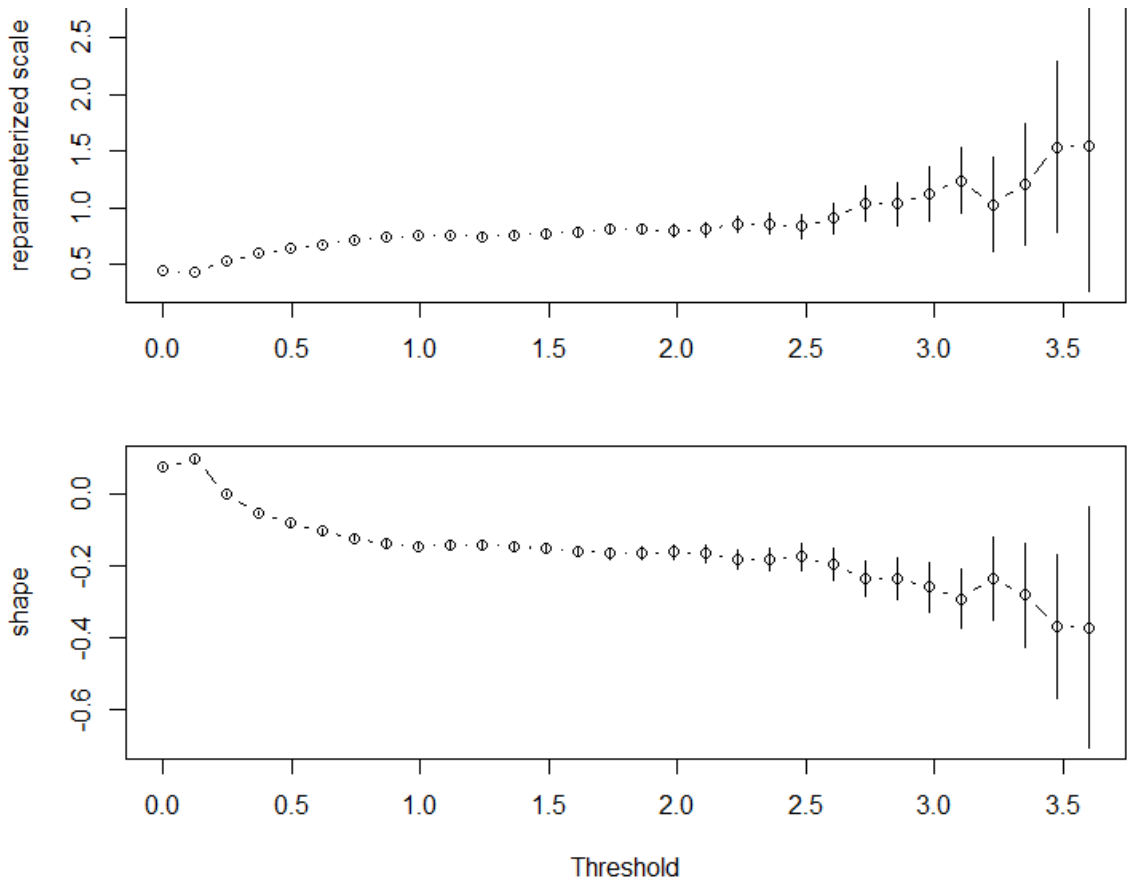
Figure 10: Parameter estimates for reparameterised scale and shape

Figure 11a shows the MRL threshold plots, where the mean excess values are linear when the threshold u is above 1. The upper and lower tails gradually underperformed for thresholds greater than 2.8 and it also shows that here the difference between upper and lower CIs (Confidence Intervals) is larger. The thresholds for scale against the threshold in Figure 11b show the same range as well.

In the last plot, Figure 11c we use Pickand's plot to find the optimal threshold. Pickand estimators are generated by a probability measure that satisfies a certain integrity condition, i.e., for a set of variables, which are independent identically distributed, $\{X_1, X_2, X_3, \dots, X_n\}$, the Pickand's estimator is based on the fact that the random variables are exponentially distributed. The expression for Pickand's estimator is as follows:

$$\widehat{\xi_{k,n}} = \frac{1}{\log(2)} \log\left(\frac{X_k - X_{2k}}{X_{2k} - X_{4k}}\right),$$

where $k = 1, 2, \dots \left[\frac{n}{4}\right]$. The Pickand's plot only estimates the threshold limits for shape. Like the above estimates, Pickand's plot (Figure 11c) shows values between 1 and 1.5 like previous estimates (Figure 10). From all these plots we can conclude that using the Graphical Method, the threshold can be $u = 1.5$.
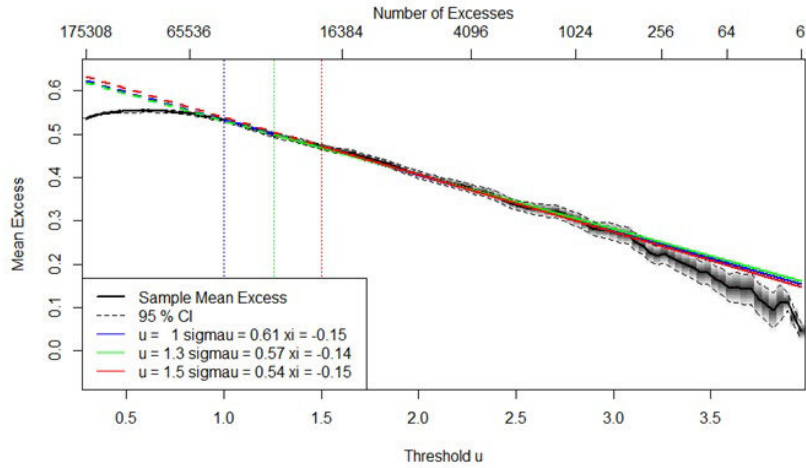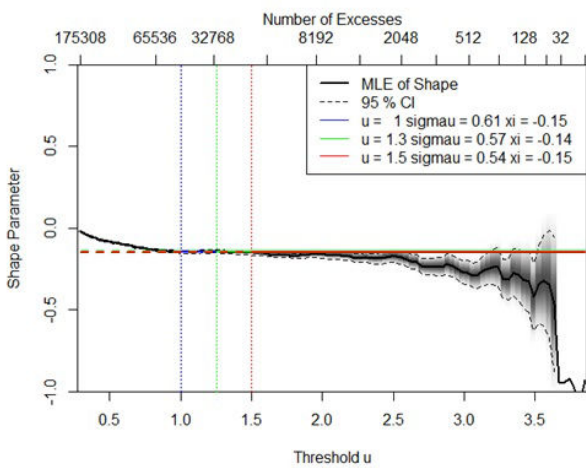
Figure 11: a) MRL plot, b) Threshold stability plot and c) Pickland's plot

2.3.3.   Threshold selection – Thompson Method

The basic model-fit method to estimate the thresholds is ideally used in the statistical analysis and has been under modifications for better assessments. The Thompson method of threshold selection plots parameter estimates of GPDs fitted using a range of thresholds against the threshold and is the basis of our automated threshold selection procedure (Thompson et al., 2009). The method plots parameters estimate of GPDs fitted using a range of threshold against the threshold via automation approach. If the threshold range is assumed as $u_1, u_2, \dots, u_n$ and σ_{u_j} and ξ_{u_j} be assumed as the maximum likelihood estimators of the scale and shape parameter based on data above the threshold $u_j = 1, \dots, n$. The scale parameter estimate can be expressed as

$$\sigma_{u_{j-1}} = \sigma_u + \xi\left(u_{j-1} - u\right) \ and \ \sigma_{u_j} = \sigma_u + \xi\left(u_j - u\right),$$

provided $u \leq u_{j-1} < u_j$. Therefore, the difference between two consecutive thresholds can be expressed as

$$\sigma_{u_j} - \sigma_{u_{j-1}} = \xi(u_j - u_{j-1}).$$

From this we can also say that Thompson method estimates can recover a known threshold from a simulated data set to a good degree of accuracy (Thompson et al., 2009).

For implementing this method, we have used the "Tea" R package (Ossberger, 2022), following the steps described in (Thompson et al., 2009), which is described as follows.

To identify a suitable threshold u, suitable values of equally spaced candidate thresholds $u_1, u_2, \ldots, u_n$, (Thompson et al., 2009), are identified. Thompson et al. (2009) suggested $n = 100$ as that gives good results. We take the first threshold, u_1 to be the median and the last one, u_n, to be the 98% quantile of the data, or the 100th value in the data set if fewer than 100 values are less than the 98% quantile. To determine which thresholds are suitable, we apply the Pearson's Chi-Square Test, a test of goodness of fit, to establish whether the observed differences between scale parameters follow a normal distribution with mean 0. If it does, the threshold u is taken to be suitable. This process is repeated until the Pearson's test indicates that the differences are consistent with a normal distribution. If it does not happen, the last threshold is returned with a warning (Thompson et al., 2009). The image below, Figure 12, shows part of the results of this analysis, in which the first column is a test number, the second column is a threshold, and the third column is the number of values above the threshold from the previous column. Next is the p-value, followed by the ForwardStop and StrongStop criteria calculated based on (G'Sell et al., 2016) and finally, the estimated scale and shape values.

```
> head(A)
  testnum threshold num.above  p.values FowardStop StrongStop est.scale    est.shape
1       1 0.2919605    175308 0.5254066  0.7452969  0.4852145 0.5474946 -0.02121727
2       2 0.3095486    168830 0.8771347  1.4209818  0.4617514 0.5542042 -0.02861513
3       3 0.3271366    162680 0.9509908  1.9525706  0.3286880 0.5606555 -0.03574234
4       4 0.3447246    156718 0.8897820  2.0157517  0.2506800 0.5675840 -0.04339411
5       5 0.3623127    151112 0.7044244  1.8563675  0.2064851 0.5737143 -0.05027572
6       6 0.3799007    145856 0.6765167  1.7350742  0.1845613 0.5787936 -0.05607471
```

Figure 12: Results of the threshold from the Thompson methodology

Using Thompson et al., 2009, approach, we need to identify when the null hypothesis of normality is not rejected ($p - value > 0.05$), so that we can choose the appropriate threshold. We can observe in Figure 13 that the p-value oscillates between values greater and less than 0.05. For consistency, we selected 0.292 as a threshold with a p-value higher than 0.05. For this threshold we have 175308 values above the threshold selected and p-value = 0.525.
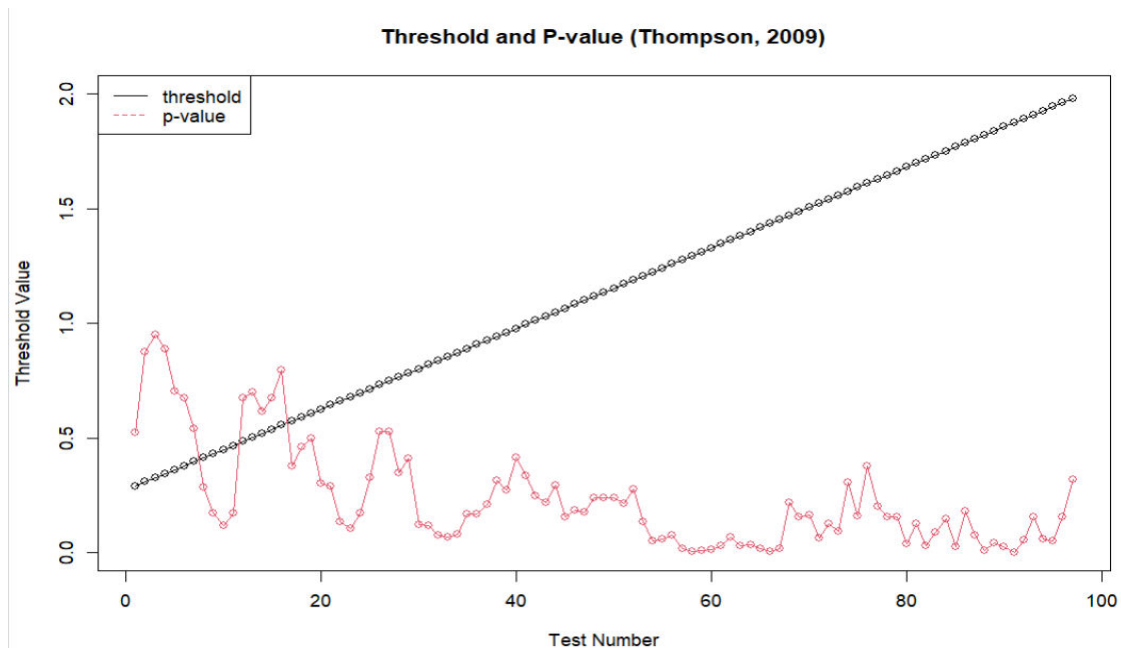

Figure 13: Results of the threshold and p-value from the Thompson methodology

2.3.4. Threshold selection – EVA package

The next methodology for finding the suitable threshold comes from the EVA R package (Bader et al., 2018), function "gpdSeqTests". This function is quite similar to (Thompson et al., 2009) methodology, but it has a more natural approach, by sequentially testing for a GPD above a certain threshold based on all data, rather than just differences in scale parameter differences. The purpose of this function is to test multiple thresholds for Goodness-of-Fit to the Generalized Pareto model using a variety of tests (Anderson Darling test, Cramer-von Mises test, Score test, etc.). For this purpose, we chose the Anderson Darling test.

In the same manner as Thompson, we take the first threshold to be the median and the last threshold to be the 98% quantile of the data, or the 100th value of the data set if fewer than 100 values are less than observed in Figure 14.

```
> head(B)
  testnum threshold num.above      p.values ForwardStop    StrongStop statistic est.scale    est.shape
1       1 0.2919605    175308 3.925898e-308  0.09509272 8.432171e-04  200.6596 0.5475803 -0.02135753
2       2 0.3095486    168830 8.879383e-276  0.09605325 1.419074e-06  180.2211 0.5541979 -0.02860712
3       3 0.3271366    162680 5.857735e-245  0.09703338 4.615812e-09  162.1227 0.5606053 -0.03567894
4       4 0.3447246    156718 6.757003e-208  0.09803373 3.411354e-11  137.5410 0.5675823 -0.04339187
5       5 0.3623127    151112 1.837488e-177  0.09905491 4.970838e-13  117.4575 0.5737123 -0.05027319
6       6 0.3799007    145856 2.198831e-155  0.10009760 1.183042e-14  102.6198 0.5787899 -0.05607014
```

Figure 14: Results of the threshold from the "gpdSeqTests" function

Figure 15 shows the threshold results in combination with the p-values results. It can be seen that p-values tend to be higher than 0.5 when the test number is higher than 40. As the p-value oscillated between higher and lower than 0.05, we chose the 10th consecutive p-value greater than 0.05 for consistency. In this case, the threshold has the value of 1.629. For this threshold, the p-value is 0.224, test number is 77 and we have 15664 values above the selected threshold.
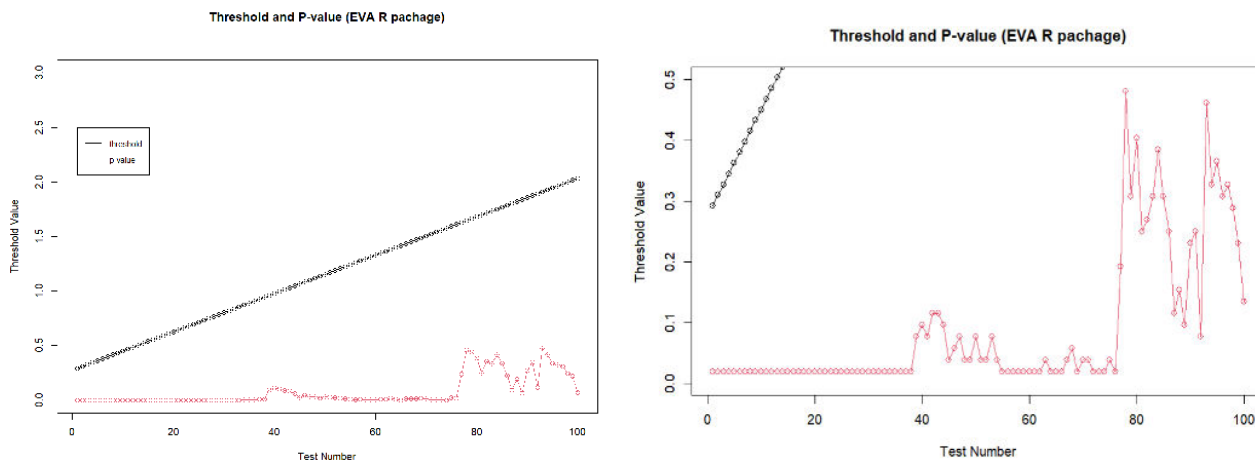
Figure 15: Results of the threshold and p-value from the Thompson methodology

2.3.5. Threshold selection – Solari Method

Solari et al. 2017 proposed another method for automatic threshold estimation using the Anderson-Darling EDF-statistic and goodness of fit test. It helps to quantify both the uncertainty of the threshold and its impact on the return period values (Solari et al., 2017), which is a drawback in the case of graphical threshold selection methods. The first step for this approach is the usual GPD approach (with function $F(x)$), however the GPD distributions are estimated using L-moments to estimate the three parameters (shape, scale, and location). Then Anderson-Darling statistic and Goodness of Fit (GOF) test with its $P - value$ are used to understand the parameter distributions.

The L-moments can handle both big- data and small sample as it uses the batch processing of extreme value data. This L-moments approach was first developed by Hosking and Wallis (1987) that estimates the three parameters (shape k, scale σ, and threshold u) of the GPD (described in section 2.3.1) from a sample size of 'n' and can be expressed as (Solari et al., 2017)

$$k = (3t_3 - 1)/(1 + t_3),$$

$$\sigma = l_2(i - k)(2 - k),$$

$$u = l_1 - \sigma/(1 - k),$$

where, $l_1 = b_0$, $l_2 = 2b_1 - b_0$, $l_3 = 6b_2 - 6b_1 + b_0$, and $t_3 = l_3/l_2$, while

$$b_r = n^{-1} \sum_{i=r+1}^{n} \frac{(i-1)(i-2)\dots(i-r)}{(n-1)(n-2)\dots(n-r)} x,$$

where, b_r is the probability weighted momentum from which b_0, b_1, b_2 are estimated for the L-moments, l_1, l_2, and l_3 and t_3 is the ratio of L-moments.

For the GOF test and its $P-value$, (Solari et al., 2017) used Anderson-Darling test (A^2), which is based on the Empirical Distribution Function (EDF) statistic proposed by (Stephens, 1974). The test, which uses the cumulative distribution function will determine if a data set comes from a specified distribution or not. For a sample size n, and z_i a cumulative distribution function $F(x)$, and for the specified order of statistics Xi; $i = 1\dots n, (x_i \le \dots \le x_n)$. The Anderson-Darling equation is as follows:

$$A^2 = \frac{1}{n} \sum_{i=1}^{n} \{(2i-1)[\log(z_i) + \log(1 - z_{(n+1-i)})]\} - n$$

For each set of the GP distribution, the parameters (shape k, scale and location u) are estimated by means of by L-moments. For the current study, the time-series data has been de-clustered using the storm window approach to identify the independent peaks. Then the threshold values are selected from the peaks and repeated values are removed from them. The A^2, and $P-values$ are calculated for each threshold by using AD statistic and L-moments. Finally, the $p-value$ and A^2 are used to select the threshold.

For the AD distribution, modified AD statistic equation by Sinclair et al., 1990 is used (known as right-tail weighted Anderson Darling statistic) as shown in the following equation. This modified one gives more weight to the upper tail distribution data than that for the lower tail (Solari et al., 2017).

$$A_R^2 = \frac{n}{2} \sum_{i=1}^{n} [(2 - \frac{2i-1}{n}) \log(1 - z_i) + 2z_i]$$

where $z_i = F(x_i)$, for the parameters of distribution F(x) from the estimate sample $\{x_{(i)}\}$. Here, the distribution of the statistics A^2 and A_R^2 depends on the sample length, the parameter estimation method, and the value of the estimated parameters, in particular the shape parameter k (Solari et al., 2017).

For the Solari method, we used two methods: one is with a fixed threshold and the other one is with the bootstrapping. The bootstrapping technique is used for estimating the thresholds (u_0 and u) and the confidence intervals of high return period quantiles. The methodology follows three steps (Solari et al., 2017):

   i)   For a series N and peaks in the series N_p the bootstrapped series N_b can be obtained from the original peaks $X_P : \{x_j, i = 1\dots., N_p$, with $j = 1, \dots N_b$;

ii)   Then, for each bootstrapped series, a threshold $u_{o,j}$, the GP parameters and high return periods are estimated using the Anderson-Darling modified approach with the threshold that minimizes one minus the p-value $(u_0 = arg \min_{u_j}(1 - p(u_j)))$;

iii)   Finally, the confidence intervals for the threshold and high return periods are estimated.

As part of this method, we used the R package available at GitHub "MultiHazard" (Jane, 2020), (the code is first used for the flood events in South Florida (Jane et al., 2020)) along with the same dataset, but instead of hourly, we used daily maximum values.

The first step to perform this analysis was to detrend the dataset used. For this analysis, we used daily maximum values from the original dataset, as Solari et al., 2017 suggested.

The wave data is detrended using the linear trend method, where a moving average of three-month period has been used to remove any seasonality from the time series. The detrended wave data can be observed in Figure 16.
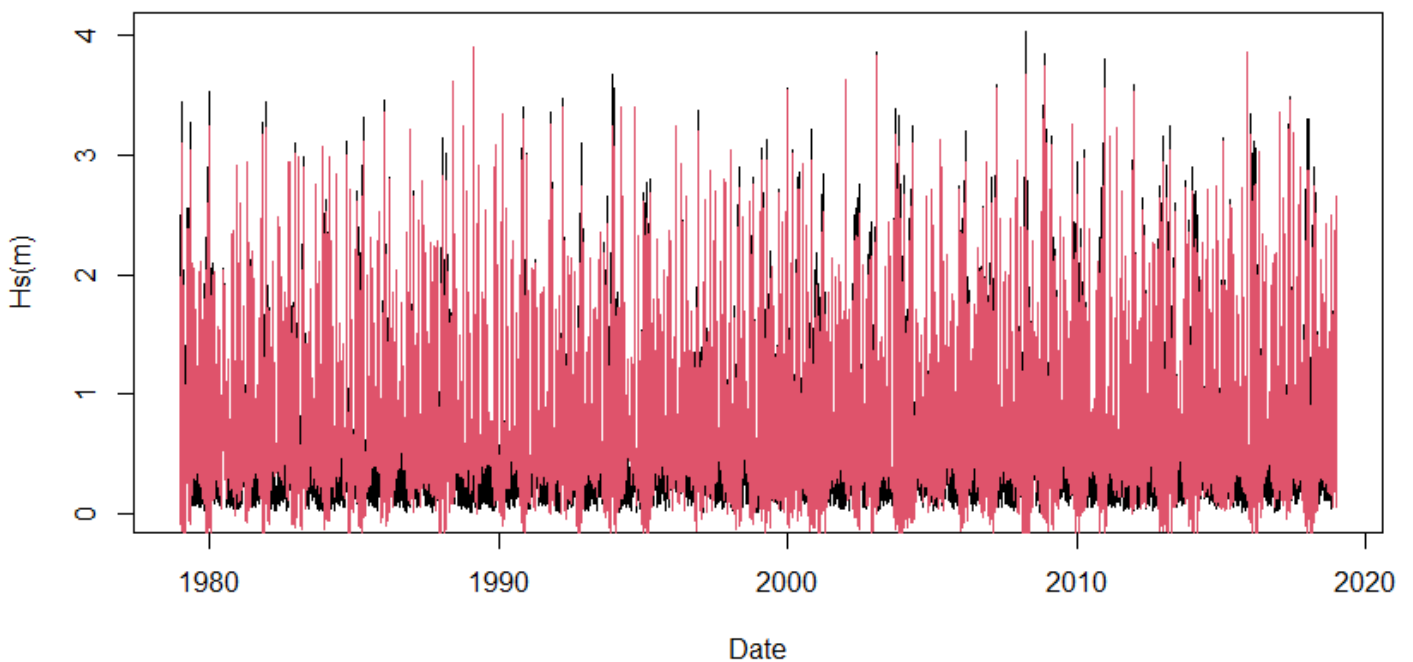
Figure 16: Detrended wave data

The next step performed was to decluster the dataset, using a specified threshold. The threshold u is specified as a quantile of the completed series, then repeated values are separated to ensure independent events. Figure 17 shows the completed, detrended record (grey circles) along with cluster maxima (red circles) identified using a 90% threshold (green line). In this plot, the threshold values range is 1.6 to 4 and while the good fit is 1.615. From these results, the number of events per year are 14.1.

The GPD fit is then used for the distribution to the observations above a threshold u, which is a quantile of the completed time-series from the Solaris AD statistics. Therefore, the GPD fit distribution uses the declustered series and detrended value to calculate the value on the original scale corresponding to u.
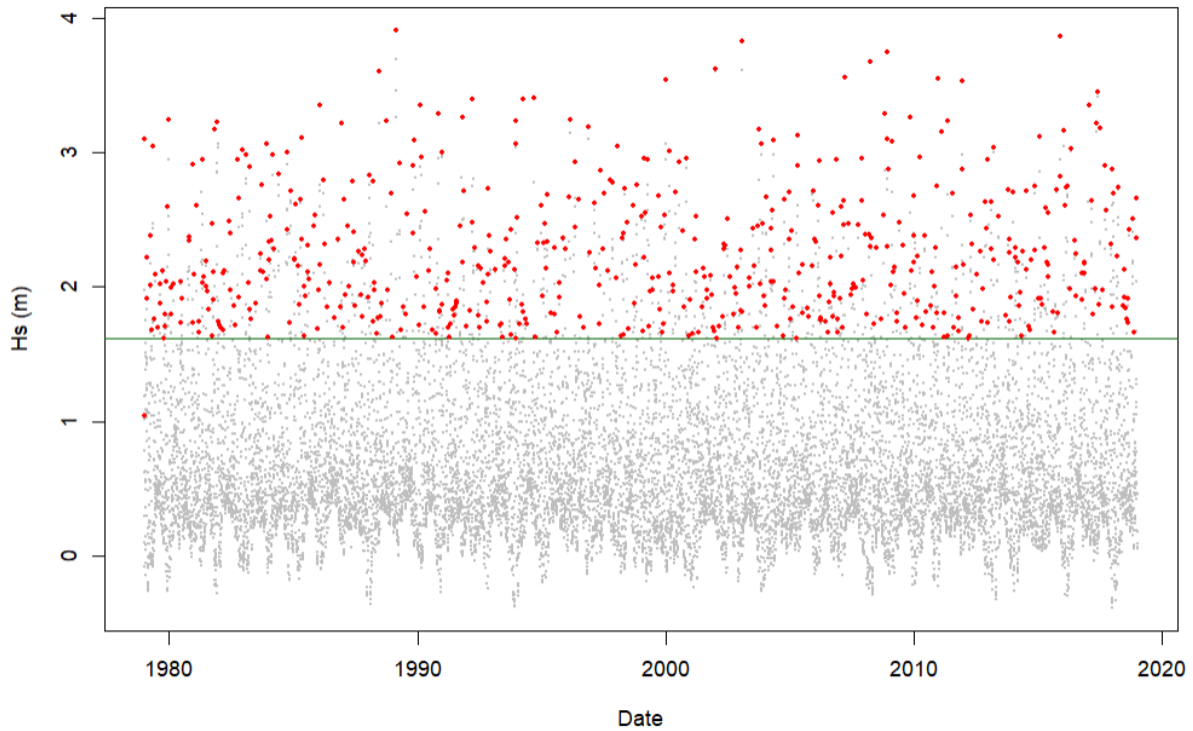
Figure 17: Declustered wave data

Table 4 shows the values for the parameter estimate for the fitted cluster maxima model. The threshold value obtained from the analysis is 2.083. The log-likelihood value of GPD is -91.354, the excess rate 0.0212 and AIC 186.708.

Table 4: Parameter estimates for the cluster maxima

Parameter	Estimate	Standard Error
Scale (σ)	0.676	0.072
Shape (ξ)	-0.315	0.049

The probability and quantile plots from the GPD diagnostics plots (Figure 18) show that the points are linearly distributed with good fit along the 0-1 straight line modelling the data. The return periods also show linearity with the data.

According to the (Solari et al., 2017) approach, we have done the first part of the analysis: declustering the time series using the storm window approach, ordering the peaks and removing any repeated values, and fit the GPD to all the peaks above each candidate threshold. The next step involves the calculation of the A_R^2 statistic and its corresponding p-value for each threshold, and the one that minimizes $(1 - p\ value)$ will be selected. Figure 19 shows the results for the A_R^2 and $p - value$, along with a set of plots for the selection of the threshold with the graphical method (MRL plot, evolution of the parameters with the threshold). This set of plots also includes 50-, 100- and 500-year return periods.

We found a threshold equal to 2.135 corresponding to $A_R^2 = 0.089$ and $1 - p\_value = 0.162,$ which is quite close to the threshold that would be selected by the graphical approach in section 2.3.2 (i.e., 1.5).

Figure 18: GPD fit diagnostic plots

Figure 19: Solaris estimated optimal threshold: a) Mean residual life plot; b) GPD shape parameter; c) Modelled GPD scale parameter; d) 50-year return period; e) 100-year return period; f) 500-year return period; g) A_R^2 statistics h) 1-P value with threshold and i) number of events per year with the threshold
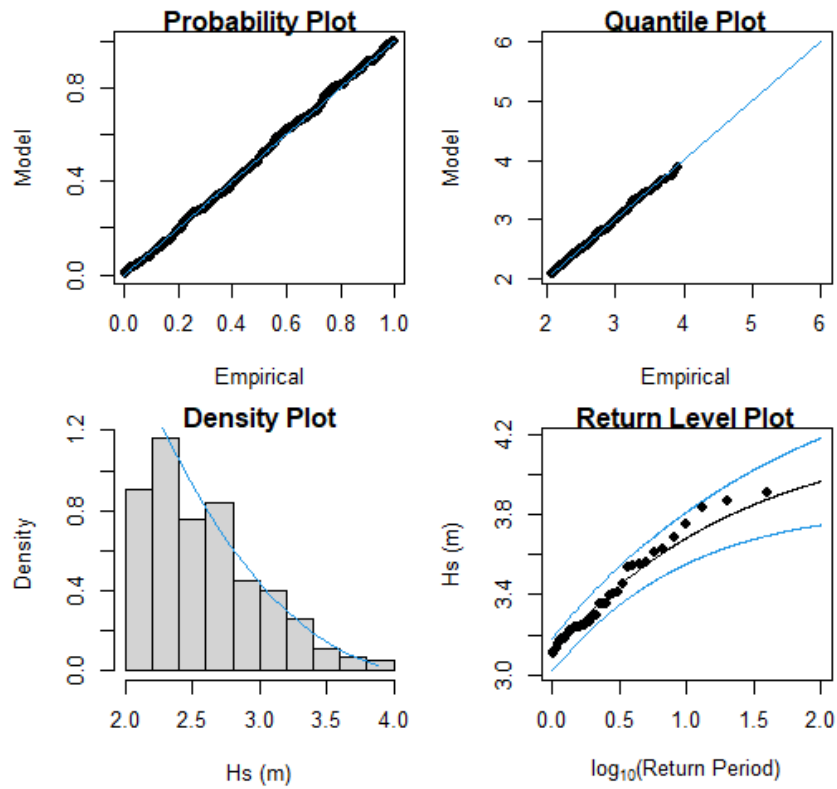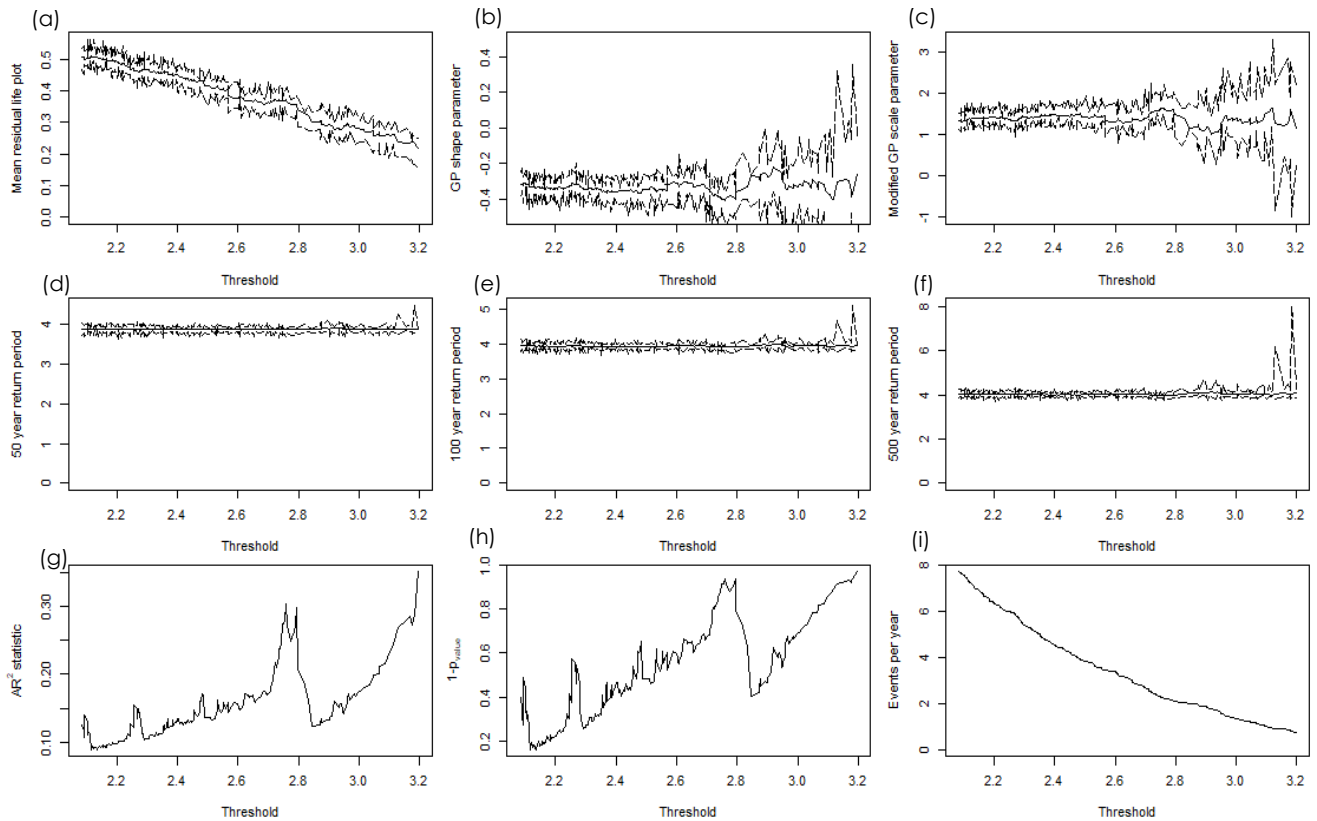
For the selected threshold we check the fit of the GPD at the selected threshold and the results are available in Figure 20. In this figure, the black continuous lines represent results obtained using a fixed threshold, while grey bars represent results obtained by estimating the threshold at every bootstrap simulation (Solari et al., 2017). The expected value obtained from data above the threshold are represented by a dashed blue line, while the green dots are represented by the extreme value distribution showing empirical distribution. Overall, these figures show a good fit to the data. The confidence intervals that were obtained with the 2 methodologies from Solari et al., 2017, tends to be similar in Figure 20a. There is a high invariance in Figure 11b and c of the confidence intervals, as we expected that histogram of the threshold to be wider when the estimation of u is automatized. This necessitates a further analysis to investigate it as Solari et al., 2017, mentioned.
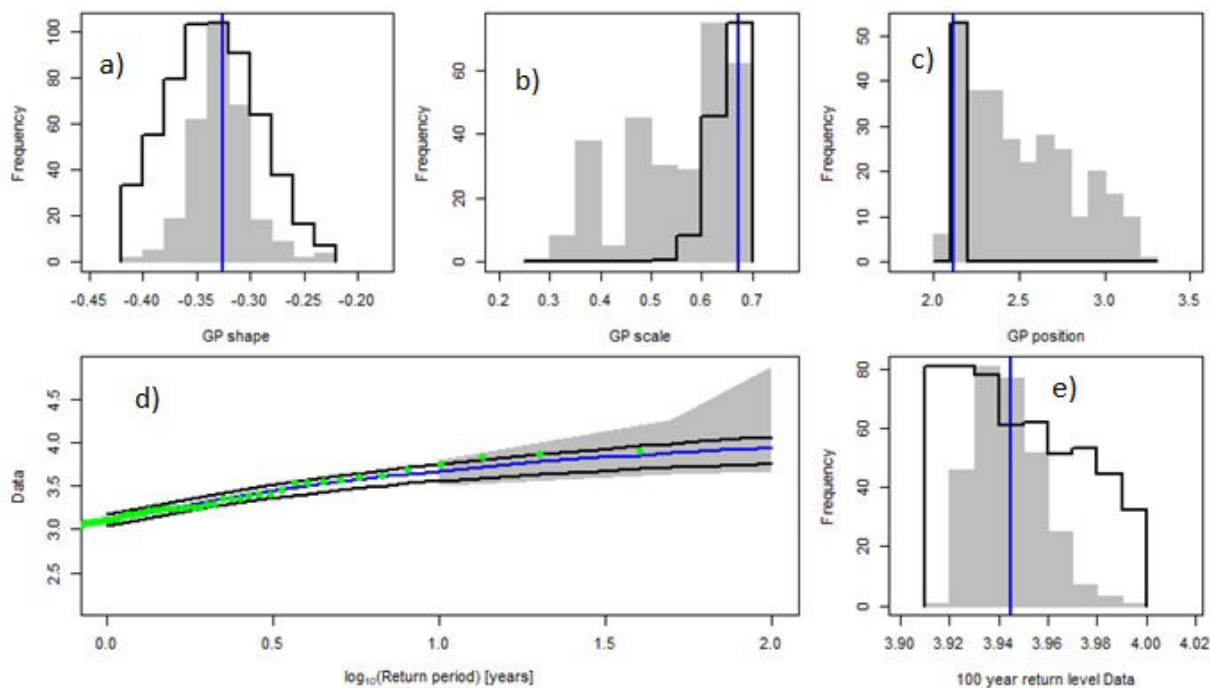
Figure 20: Histograms obtained via nonparametric bootstrapping simulations for a) Shape; b) Scale; c) Position; d) 100-year return levels simulations: Data are $Hs(m)$; e) Extreme value distribution: 100 -year return level of Data - $Hs(m)$.

The results of the parameters for Solari method and the return levels are presented in Table 5. We can see that the return levels of $H_s(m)$ tends to increase from 3.113m to 3.945m from 1 to 100 years.

Table 5: Parameter estimates and return levels $Hs(m)$ – Solari Method

Parameter	Value	Parameter	Value
Scale (σ)	-0.325	RL 30-years	3.830
Shape (ξ)	0.672	RL 40-years	3.861
Threshold (u)	2.117	RL 50-years	3.884
RL 1-year	3.113	RL 70-years	3.915
RL 5-years	3.550	RL 90-years	3.936
RL 10-years	3.678	RL 100-years	3.945
RL 20-years	3.780		

2.3.6. Results on Threshold selection and Return level

The selection of an appropriate threshold in extreme value analysis is a crucial step and thus, several methods have been proposed in the scientific literature. In this report, the threshold selection was carried out based on the Graphical method, Thompson (Thompson et al., 2009), Anderson-Darlin test (Bader et al., 2018), and Solari method (Solari et al., 2017) in order to find the most suitable one for the given dataset. After a comprehensive evaluation, the Anderson-Darlin test was found to be the best suited for this dataset, as the Thompson method had a low threshold, while the Solari results gave a considerably higher value. The results of the Anderson-Darlin test were then used to obtain the optimal threshold that could provide the results for calculating the return levels for the significant wave heights. The results of the four methods, together with their return levels are presented in **Table 6**.

Table 6: Results on threshold selection

Threshold selection
File name: hs_eval_massa_05.nc

Method	Threshold	20yRL	50yRL	100yRL	200yRL
Graphical	1.5	3.105	3.41	3.588	3.732
Thompson	0.292	3.17	3.742	4.145	4.524
Anderson-Darling test	1.629	3.115	3.417	3.591	3.728
Solari	2.117	3.78	3.884	3.945	3.993

Prior to fitting the distribution with the selected threshold ($u = 1.629$), the dataset was declustered to ensure that any trends had been removed, Figure 21.
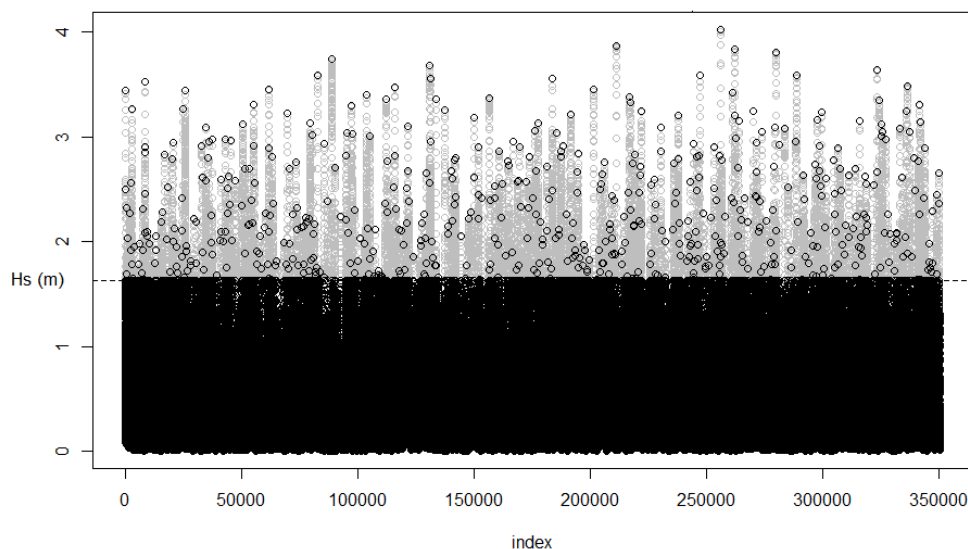
Figure 21: Declustered data with the selected threshold

The threshold that was selected allowed us to calculate the return levels for different distributions. These distributions included the GPD fit, the exponential and the Point Process Approach (Section 2.4). In order to ensure the accurate selection of the suitable distribution for the dataset, all of these distribution results will be further investigated in Section 2.5. Through detailed analysis in this section, we aim to determine the best distribution that suits the dataset and will provide accurate results.

The results of the GPD fit using the estimation method MLE and the exponential distribution are provided in **Table 7**, along with the empirical and model quantiles from Figure 22 and Figure 23. In the GPD fit, both quantile plots show a good fit, whereas the exponential fit results are overestimated. The AIC and BIC results for the GPD fit are also better than the exponential fit.

Table 7: Parameter estimates of the GPD fit and exponential distribution using the selected threshold

Distribution	Parameter	Estimate	Value	Results
GPD Fit	Scale (σ)	0.860	AIC	624.385
	Shape (ξ)	-0.324	BIC	633.148
	Negative log likelihood	310.192		
Exponential	Scale (σ)	0.646	AIC	667.060
	Shape (ξ)	NA	BIC	671.442
	Negative log likelihood	332.530		
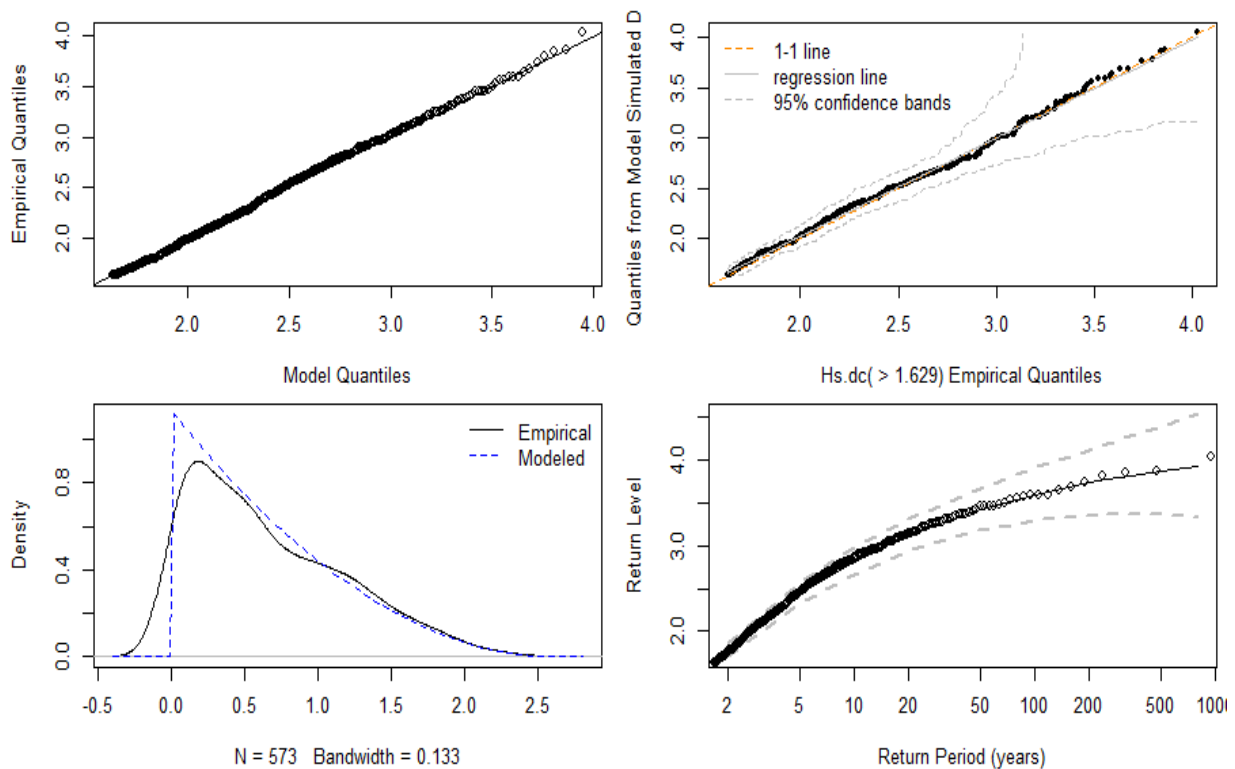
Figure 22: Result plots of the GPD fit
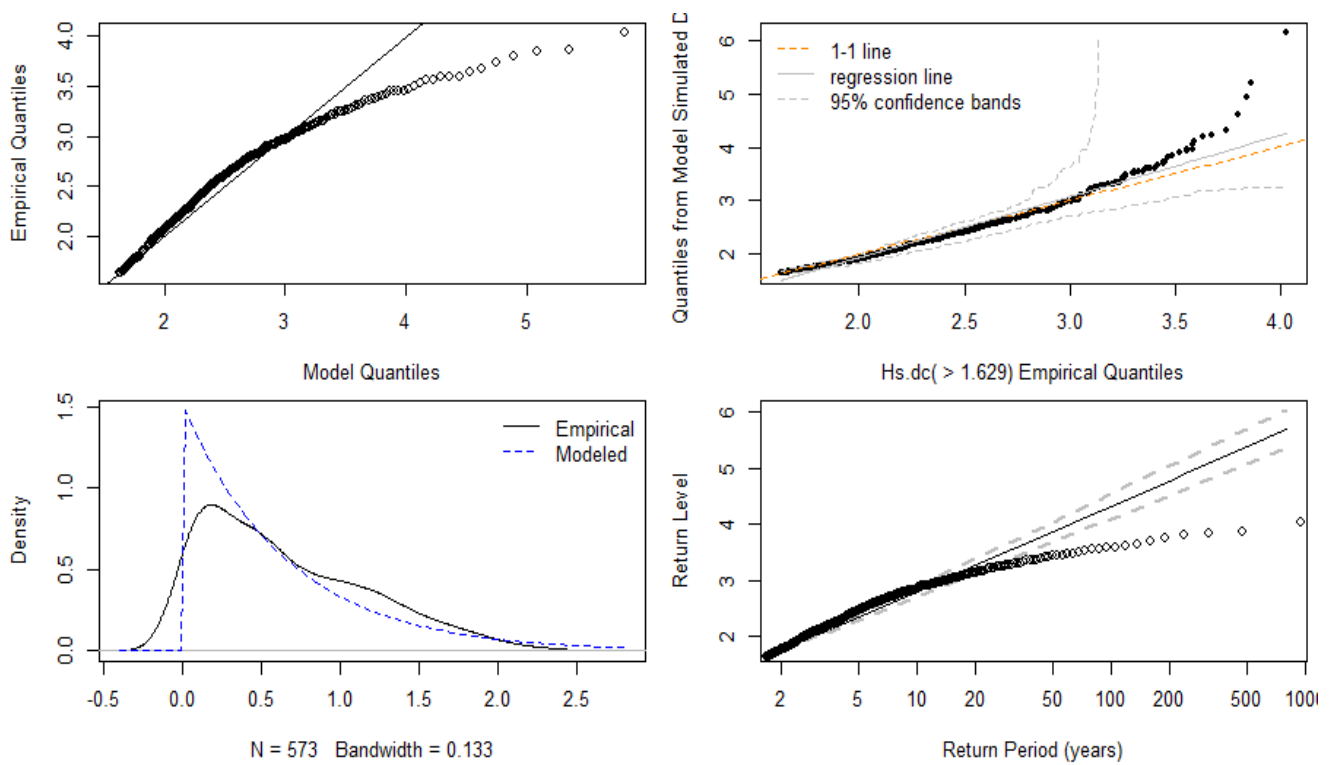
Figure 23: Result plots of the exponential distribution

In the following table, we calculated the return levels for both distributions. According to this table, we can see that it is important to choose the best distribution since the results are quite different; for example, the RL (Return Level) for 50-years differs by approximately 0.4m in terms of $H_s(m)$.

Table 8: Return levels of $Hs(m)$ with the Confidence Intervals (CI) calculated with the selected threshold

Distribution		95% lower CI	Estimate	95% upper CI
GPD fit	RL 25-years	3.002	3.188	3.374
	RL 50-years	3.179	3.408	3.637
	RL 100-years	3.300	3.584	3.867
	RL 200-years	3.372	3.724	4.076
Exponential	RL 25-years	3.252	3.394	3.537
	RL 50-years	3.664	3.842	4.020
	RL 100-years	4.075	4.290	4.504
	RL 200-years	4.487	4.737	4.988

2.4. Point Process approach

The point process (PP) modelling approach is considered as an elegant alternative of extreme value analysis, due to its capability in modelling both the intensity rates and frequency of extremes occurrence. The PP analyses extremes by unifying and extending the EVT based on the BM and POT approaches. With respect to the POT approach, the PP process is more associated with the variations in the excesses. More in detail, the PP describes occurrences in space or time and be used to model excesses over the threshold as occurrences in time.

The block maxima $M_n = max\{X_1, X_2, X_3, \ldots, X_n\}$ should be distributed with the GEV for the normalized constants $a_n > 0$ and b_n, for the occurrence of events (Y_i). As described in Boano-Danquah et al., 2020, this leads to the sequential point process T_n on R^2 given by

$$T_n = \left\{ \left(\frac{i}{(n+1)}, \frac{(Y_i - b_n)}{a_n} \right), i = 1, \ldots, n \right\}$$

The data are made stationary by a time-varying threshold which has a positive shift factor. The asymptotic theory of threshold exceedances, under suitable normalisation, shows that the exceedances over threshold are fitted to a non-homogeneous Poisson process. Further mathematical details on PP can be found in (Karr, 2017).

The selection of the threshold is a compromise between the variance and the bias. On the one hand, if the threshold is too small, this goes against the asymptotic properties which underline the GPD deviation. On the other hand, if the threshold is too high, the excesses over the threshold are too small to estimate scale and shape parameters, this leads to high variance. This means that the selection of the threshold is crucial to understand whether the model provides a good approximation versus the parameter estimates' variance (Boano-Danquah et al., 2020).

Maximum likelihood estimation (MLE) is obtained for a GPD model, and the resulting parameter estimated are converted to those of an approximately equivalent PP model. For PP, three initial value estimates are calculated along with their associated likelihood values, **Table 9**.

Table 9: Parameter estimates of the PP using the selected threshold

Parameter	Estimate	Value	Results
Location (μ)	1.177	AIC	2381.712
Scale (σ)	1.007	BIC	2394.858
Shape (ξ)	-0.324		
Negative log likelihood	1187.856		

Empirical and model quantiles are provided in the plots in Figure 24, along with the equivalent GEV at different return periods and levels. From the density plot we can observe underestimated data, although the QQ plots tends to have a good fit.
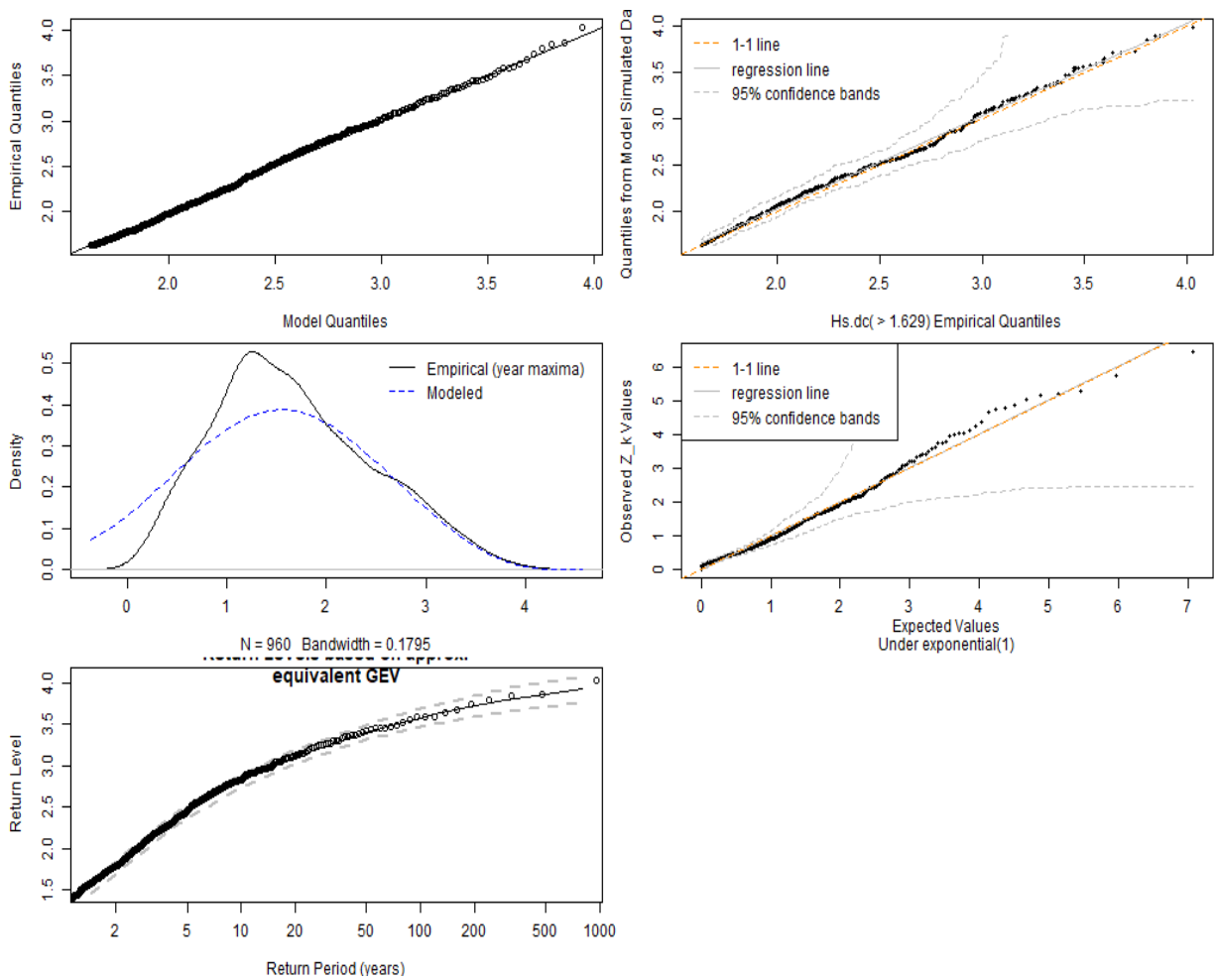
Figure 24: Result plots of the PP

Table 10 presents the return levels for the PP distribution. The $Hs(m)$ look very close to the estimates for the GPD and Exponential fit in Table 8.

Table 10: Return levels of Hs (m) with the CI calculated with the selected threshold for the PP

	95% lower CI	Estimate	95% upper CI
RL 25-years	3.096	3.181	3.265
RL 50-years	3.313	3.405	3.497
RL 100-years	3.477	3.583	3.688
RL 200-years	3.600	3.724	3.847

2.5. Summary results

In the table below, **Table 11**, we summarise the results of the EVA for the POT against the BM methods in the case of historical time series (1979-2018) of significant wave heights (H_s (m)) in Marina di Massa. Note that the POT method considers hourly data (except for the Solari method, which considers daily maxima), whereas the BM method considers annual maxima.

For each method, location-scale and shape parameters and accuracy in terms of AIC and BIC is showcased. **Table 11** gives the return level at different return times and the related confidence intervals are provided. For POT methods, GPD performs better in terms of goodness-of-fit (lower prediction error). Here, the selected threshold is 1.63 (see section 2.3.6 for details). In the case of BM, the GEV shows lower AIC and BIC values compared to Gumbel.

As expected, in the POT method, the Weibull distribution gives the best results (light tail), which is why the exponential method is not appropriate for this dataset since the shape parameter is negative. The GEV distribution performs better in terms of negative log likelihood, AIC, and BIC compared to the PP distribution. This particular dataset with significant wave heights is therefore best suited to the POT/GPD distribution.

Using the selected threshold, return level against return period have been plotted (Figure 25), along with histogram (Figure 26), density (Figure 27), quantile-quantile (QQ) plot of the data quantiles against the fitted model quantiles (Figure 28) and QQ plot of quantiles from model-simulated data against the data plots (Figure 29) across all methods.

QQ plots reveal that distributions of observations (empirical quantiles) and simulated data in the GP and PP methods (POT) are very similar. This is still valid for GEV in the BM method. The exponential (POT) and Gumbel (BM) show significant deviations at the high end of range i.e., outliers in the upper right corner.

In regard to the histograms (Figure 26) and density plots (Figure 27), the GPD method shows good results. As for the GEV method, the density plot shows good results, but the histogram shows an underestimate.

The last 2 series of plots, QQ plots (Figure 28 and Figure 29) demonstrate a skewness to the right (exponential and Gumbel) and left (GEV), while GPD and PP tend to show a normal distribution. On the Gumbel and GEV QQ plots, the data appears to be peaked in the middle, which is also evident in the histograms. Looking at the QQ2 plots for the GP and PP, the model for GPD appears to be better fit.

We can conclude from these plots and results that we have achieved the best results for the **POT/GPD method**, and we should retain the return levels results for this method only.

Table 11: Results for the POT and BM

Return period for Hs (m)

File name: hs_eval_massa_05.nc

Method	Peak Over Threshold			Block Maxima	
	GPD	**Exponential**	**PP**	**GEV**	**Gumbel**
Sample size	350616	350616	350616	40	40
Location	NA	NA	1.131	3.209	3.156
Scale	0.884	0.657	1.051	0.324	0.309
Shape	-0.337	NA	-0.337	-0.316	NA
Negative log-likehood	309.125	332.529	1177.781	10.774	13.858
AIC	622.25	667.058	2361.562	27.548	31.716
BIC	630.952	671.409	2374.615	32.614	35.094
RL_25years	3.197	3.406	3.19	3.861	4.143
CI_25y_Lower	3	3.26	3.105	3.727	3.864
CI_25y_Upper	3.394	3.551	3.275	3.996	4.421
RL_50years	3.417	3.861	3.414	3.936	4.36
CI_50y_Lower	3.173	3.678	3.323	3.779	4.034
CI_50y_Upper	3.66	4.044	3.505	4.092	4.686
RL_100years	3.591	4.317	3.589	3.995	4.576
CI_100y_Lower	3.287	4.097	3.487	3.811	4.202
CI_100y_Upper	3.895	4.537	3.692	4.178	4.949
RL_200years	3.728	4.772	3.728	4.042	4.79
CI_200y_Lower	3.347	4.515	3.609	3.83	4.369
CI_200y_Upper	4.109	5.03	3.847	4.255	5.212

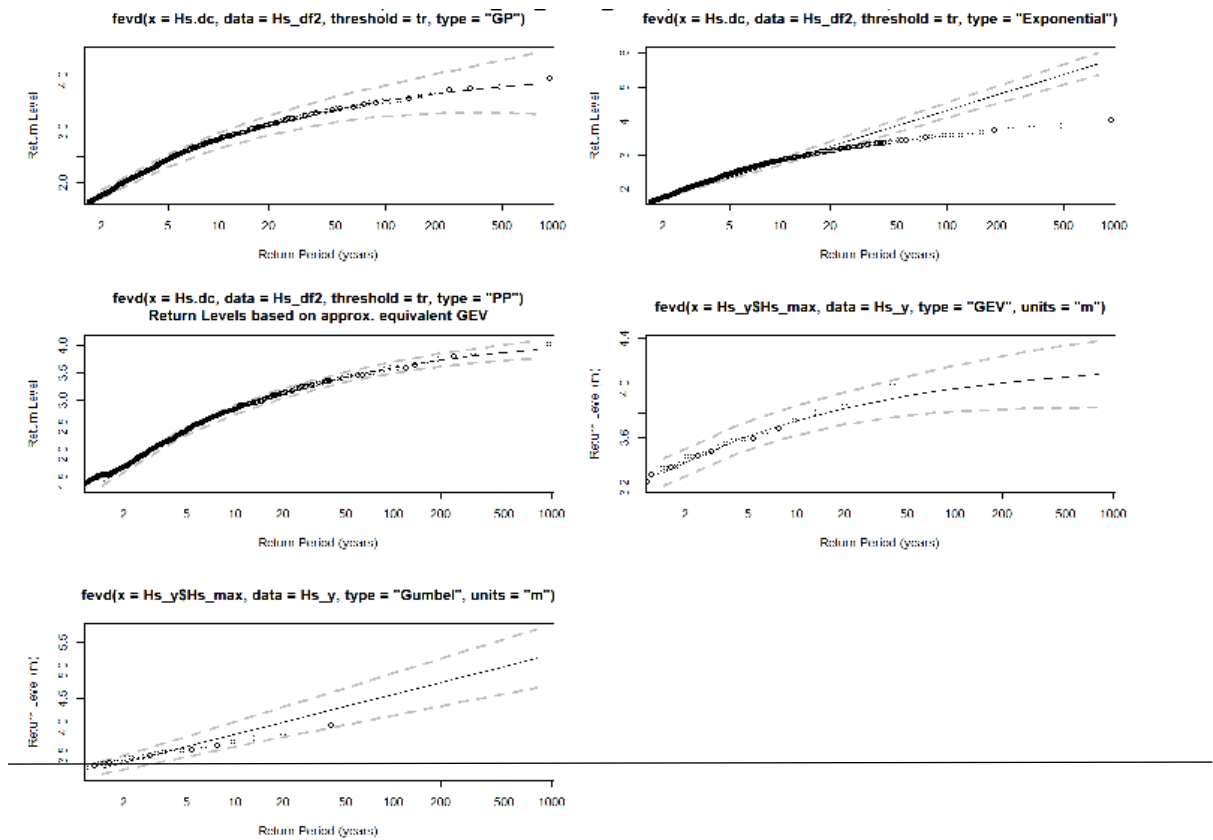POT includes a threshold= 1.63 and uses hourly data; the Block Maxima uses annual maxima data.

Figure 25: Return level calculated using the POT and BM methods
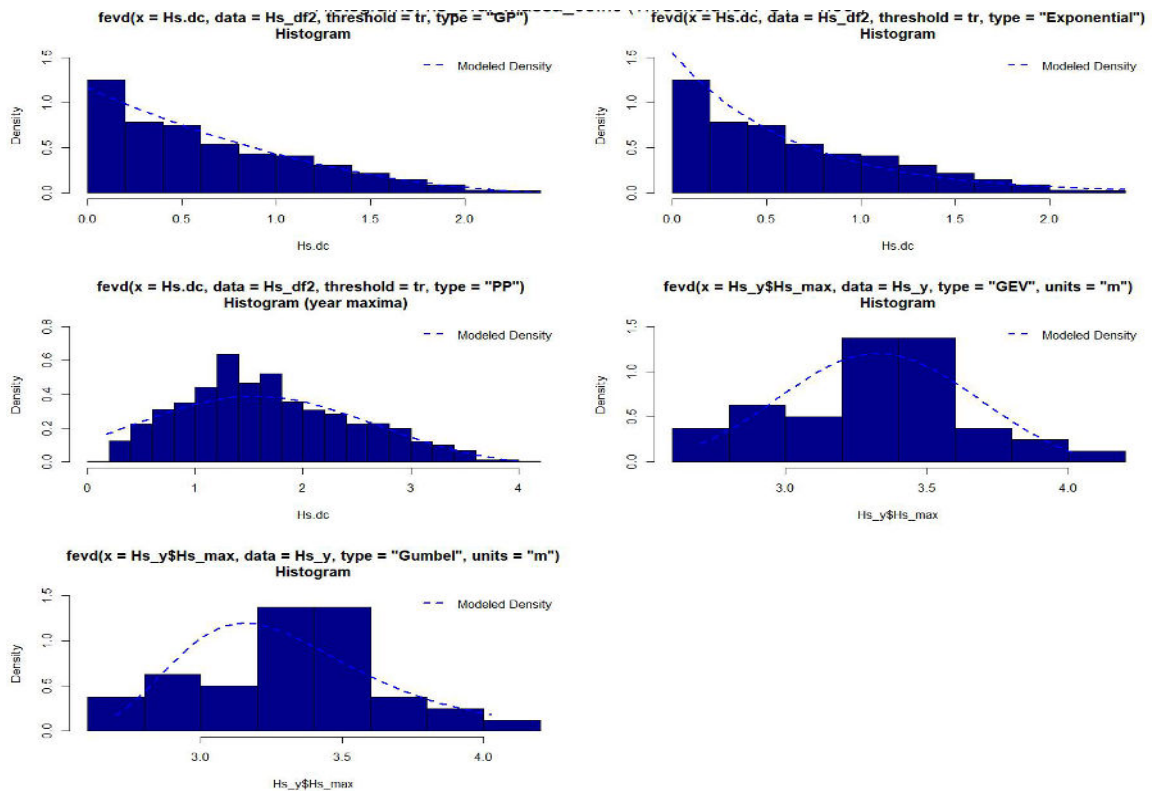

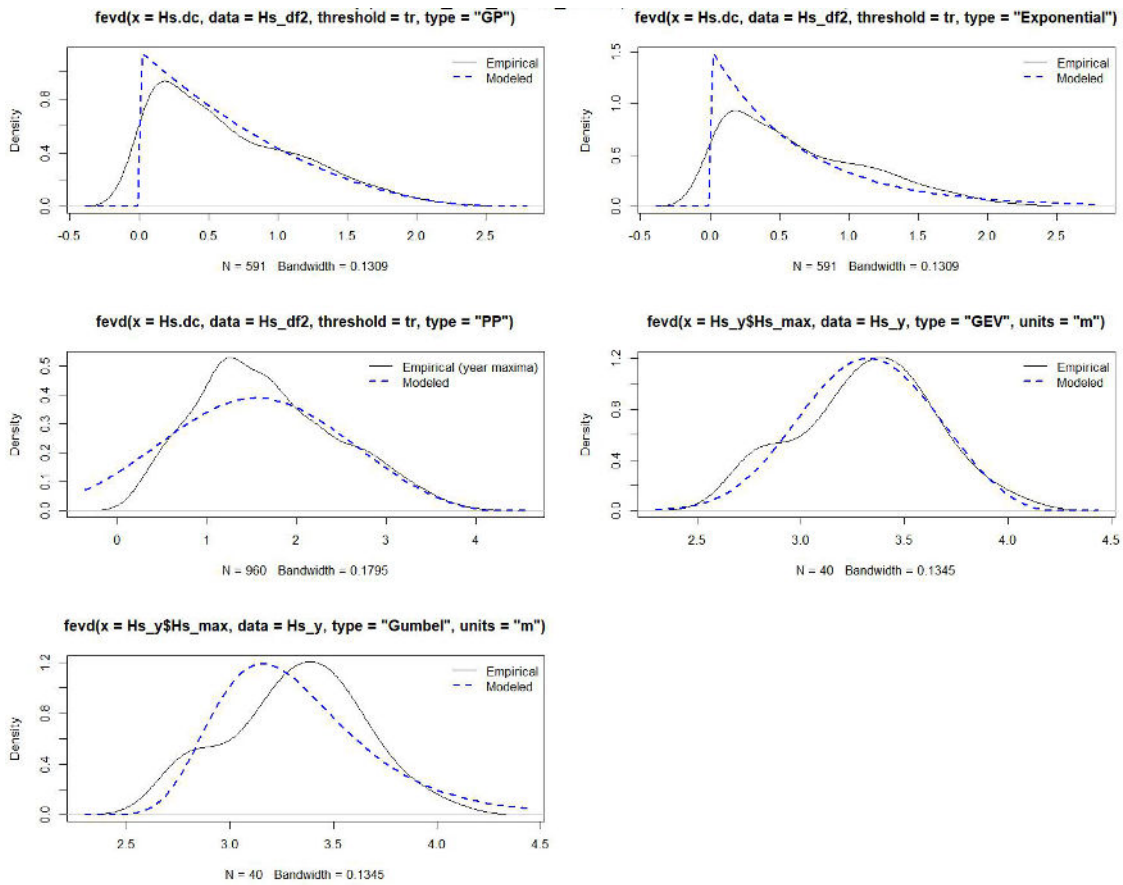*Figure 26: Result plots – Histograms - for the POT and BM methods*

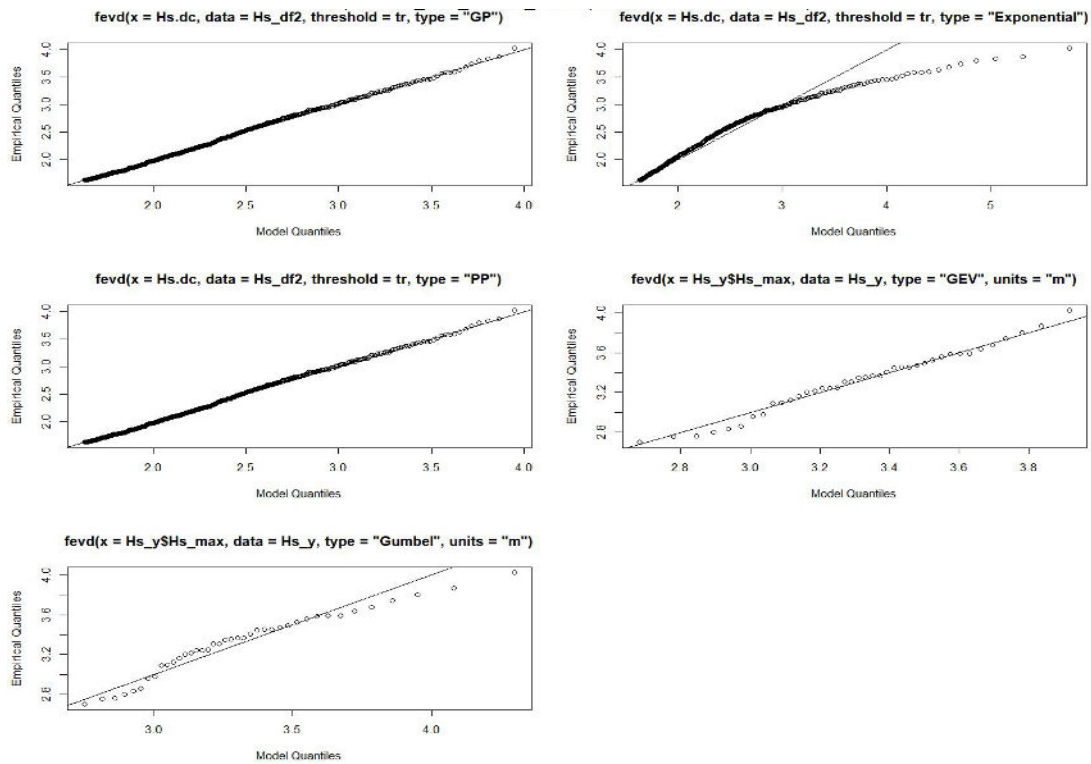*Figure 27: Result plots – Density plots - for the POT and BM methods*

Figure 28: Result plots – QQ plots - for the POT and BM methods
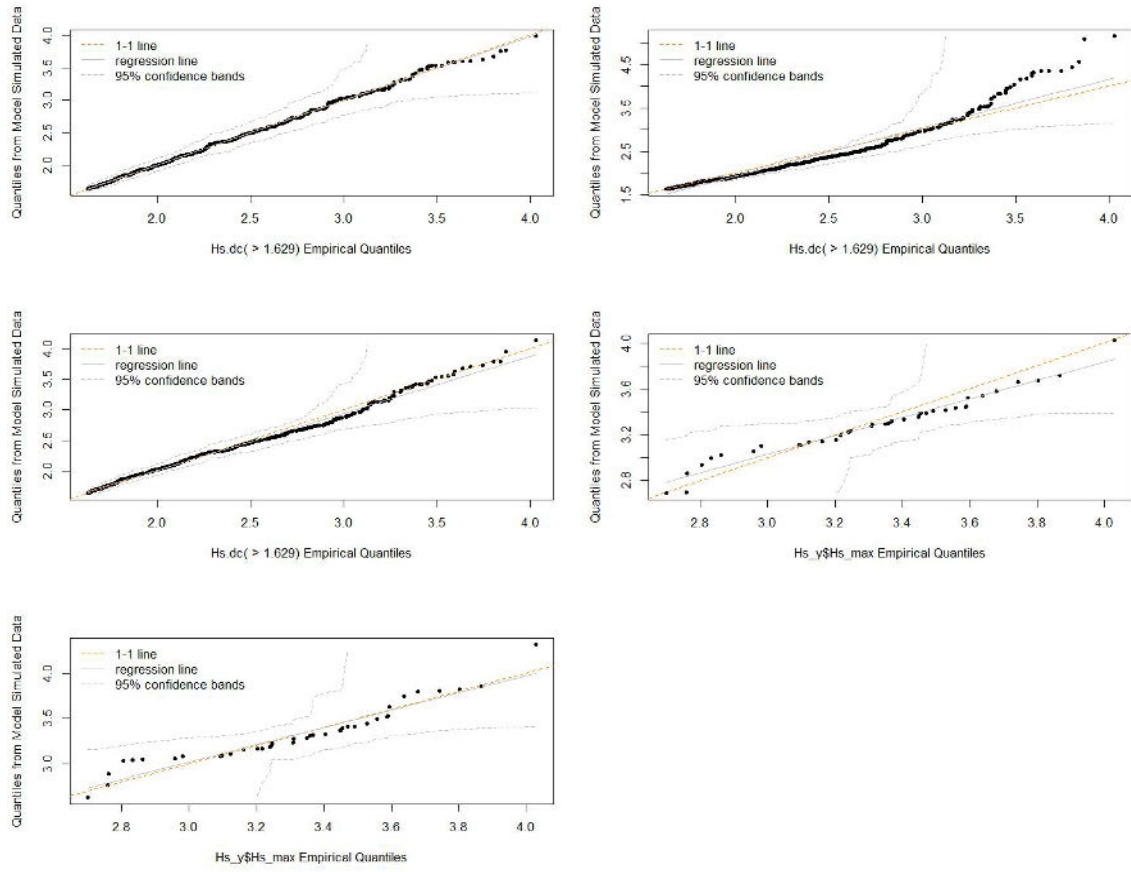
Figure 29: Result plots – QQ2 plots - for the POT and BM methods

3.DATA CLUSTERING

Hydrogeomorphological disasters can be considered as stochastic point processes acting in both temporal and spatial dimensions (Stoyan, 2006). Point process can be analyzed in terms of their random dispersion, distribution and clustering. Cluster analysis is an unsupervised technique that aims to detect subgroups among the observations that, based on certain features characteristics, are similar to each other. In this framework, data clustering aims at selecting and grouping homogeneous conditions, i.e., the study of typical coastal flooding conditions and other hazards. Data clustering can be used in the analysis of circulation types classification.

The partitioning clustering also known as the centroid-based method, is a type of clustering that divides the data into non-hierarchical groups. One of the most used unsupervised machine learning algorithms in this framework is the k-means clustering proposed by (MacQueen, 1967). k refers to the number of clusters centres (which corresponds to the mean of points assigned to the cluster) exploited to classify the data sets. The algorithm classifies objects in multiple groups i.e., k similar clusters of data points, so that the total intra-cluster variation (known as total within-cluster variation) is minimized. As an example, in the case of inundation depths, inundation depths with specific features are identified and the data points with minimum Euclidean distance from the cluster center are identified as control points for each cluster (Lin et al., 2013). There are several k-means algorithms available. The standard algorithm defines the total within-cluster variation as the sum of squared Euclidean distances between points and the corresponding centroid (Hartigan & Wong, 1979):

$$minimize \sum_{i=1}^{k} \sum_{x_j \epsilon S_i} ||x_j - \mu_i||,$$

here x_j is the jth data value, k is the number of clusters and μ_i is the cluster centre (mean) of data included in cluster S_j.

The k-means algorithm starts with an initial group of randomly selected centroids, which are used as the beginning points for every cluster, and then performs iterative calculations to optimize the positions of the centroids. It stops optimizing clusters when either all the centroids are stabilized and there is no change in their values, or the number of iterations has been achieved.

The determination of the right number of clusters is not trivial. To do this, many tests and methods have been introduced (Grover & Vriens, 2006). Possible approaches are the within-cluster sum of squares or visual representation of the data in dendrograms or scatterplots.

The elbow method (within-sum-squares) accounts for the percentage of explained variance as a function of the number of clusters. At some point adding clusters will not add much information; the best k is after the variation isn't very steep, giving an angle (elbow) in the graph (Thorndike, 1953).

The hierarchical technique avoids this problem altogether (Johnson, 1967). The hierarchical clustering technique is divided into agglomerative and divisive. In the agglomerative technique, which is the most used, each point is initially considered as an individual cluster. At each iteration, the two closest clusters are merged, and the proximity matrix is updated until a single cluster remains. Briefly, all data are initially considered as a single cluster and the data points which are not similar are separated at each iteration.

In the following plots, significant yearly maxima of wave heights H_s (m) against river discharge yearly maxima $Q_{max}(m^3/s)$ have been considered. Note that here we provide just an example of application with the available variables Hs_max and Q_max. This does not mean that these are the optimal combination of variables to be analyzed with this approach. As such, some difficulties in the interpretation of the results could arise.

Clustering algorithms form groupings in such a way that data within a cluster have a higher measure of similarity than data in different clusters. Similarity measures are distance measures including e.g., correlation, Euclidean, probabilistic, cosine distance. The distance is the length of the vector between the two points, called the Euclidean distance. The lower the distance, the closer the observations are in the cluster. The higher the similarity, the more similar the observations are in the cluster. To compute distance matrices, either Euclidean or correlation-based measures can be used. In general, it is recommended to standardize the variables before distance matrix computation. Standardization makes variables comparable among different scales.

We implemented a k-means clustering which implicitly assumes Euclidean distance. In the figure below, Figure 30, based on the elbow method output, we selected 2 clusters and ran the algorithm n times with random initialized centroids. By minimizing the model's quality measurements, the best solution should be reached. This measurement is the total within-cluster sum of squares (WCSS).

To perform cluster analysis, the *factoextraR* (Alboukadel Kassambara, 2017) and *cluster* (Martin Maechler, 2022) R packages have been used.
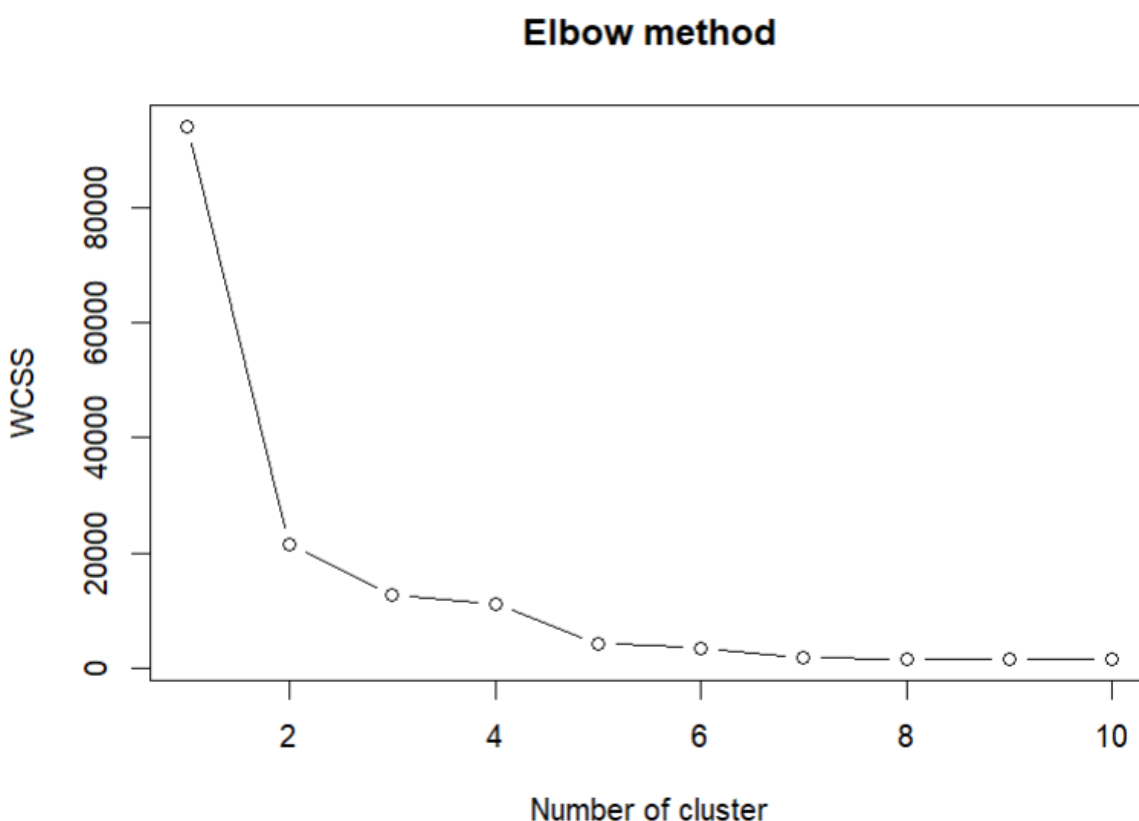
Elbow method

Figure 30: Elbow method – numbers of clusters and within-cluster sum of squares (WCSS)

In the resulting plot, data observations are represented by points, using the features, if the number of variables is greater than 2. A concentration ellipse is drawn around each cluster. In Figure 31, the first two features are Hs_max (dim1, on the x-axis) and Q_max (dim2, on the y-axis). To find the mean of the variables in each cluster, the *aggregate* function (Becker et al., 1988). We can also exploit parallel coordinates to understand what clusters represent by looking at how the variables compare across the clusters. Based on Figure 31 two groupings can be identified:

- Blue sample (cluster 1), with widespread (low to high) Hs_max and Q_max;
- Red sample (cluster 2), with low Hs_max and low Q_max;

Figure 31: Cluster plot

The two components capture around 39% and 33% of the total variance. The hierarchical clustering (also known as clustering of clustering) starts with a distance matrix (in this case, the Euclidean dist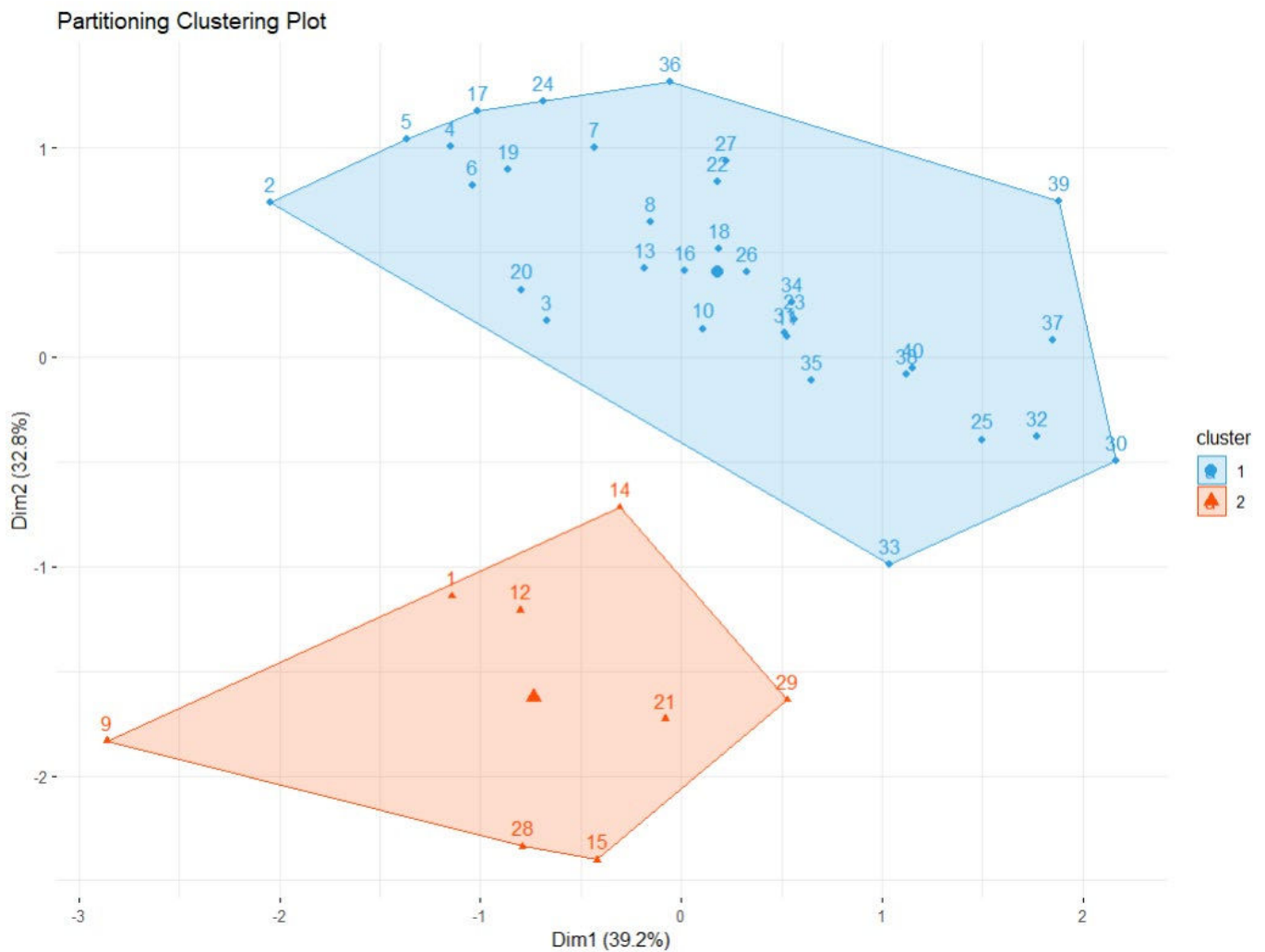ance has been adopted as the default method). The default output is a dendrogram. The dendrogram is a tree-like structure with data points on the x-axis, while the cluster distance lies on the y-axis. Going down in a path, the clusters become smaller and smaller until the level reaches the data sample. In the hierarchical clustering, no assumption on the number of clusters is made. Slicing the structure horizontally, all the branches below that line are an individual cluster at the highest level. This means that, once created the clusters, the relationship between the subsequent subclusters that can be formed is still known and the level of granularity of clustering can be increased or decreased. At every branching, all the data having the membership at each level belongs to a certain class. To infer the class entity, it is necessary to go through few individual samples of each level in the cluster to find out what feature is common in the resulting cluster. To understand and visualize graphically how the dendrogram will be cut into clusters, the number of clusters should be specified. This dendrogram, Figure 32, was realized using a partition of 2 clusters. The first cluster (far left) is composed of 9 observations. The second cluster to the right has 31 data. Cutting the dendrogram higher, we would have fewer final clusters but with a lower similarity.
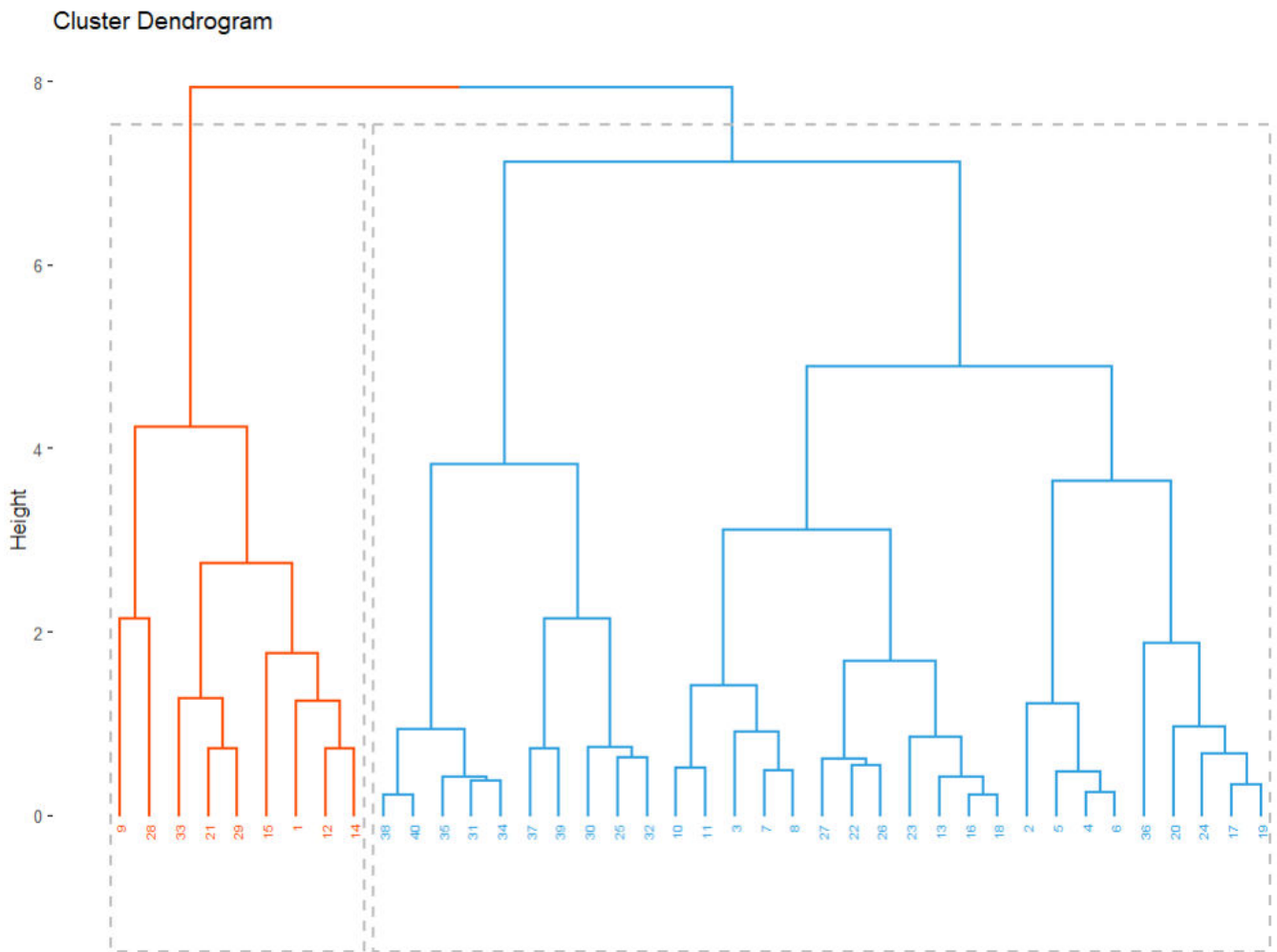
Cluster Dendrogram

Figure 32: Cluster Dendrogram

There is no incorrect or correct number of clusters. The clustering outcome can be measured based on compactness (how much observations in each cluster are similar to each other) and closeness (i.e., how much observation in a cluster are closer to each other than they are to other data in other clusters).

To validate clusters, we can exploit the silhouette (Si) analysis, which is aimed to calculate the similarity of each observation to the assigned cluster (Rousseeuw, 1987). The silhouette width is defined as follows:

$$\frac{(b - a)}{\max(a, b)}$$

where a is the average dissimilarity or intra-custer distance (i.e., an average measure of how much the data samples are distinct, opposite of the similarity) between the ith point x and the other points in the cluster to which x_i belongs to. The average distance between all cluster (inter-cluster distance) is denoted by b.

The metric (i.e., the silhouette width) ranges from −1 to 1 for each observation and can be interpreted as follows:

- Observations with a negative $S(i)$ are probably assigned to the wrong cluster;

- A small $S(i)$ (around 0) is indicative of the fact that the observation lies between two clusters;

- Observations with a large silhouette $S(i)$ (almost 1) should be very well matched to the assigned cluster.

Data in Figure 33 are ordered within each cluster from the highest to the lowest $S(i)$. As can be seen, our two clusters solution returns an average silhouette coefficient of 0.31. One of the objects in cluster 1 has a slightly negative silhouette coefficient indicating it may be in the wrong cluster (observation 33).

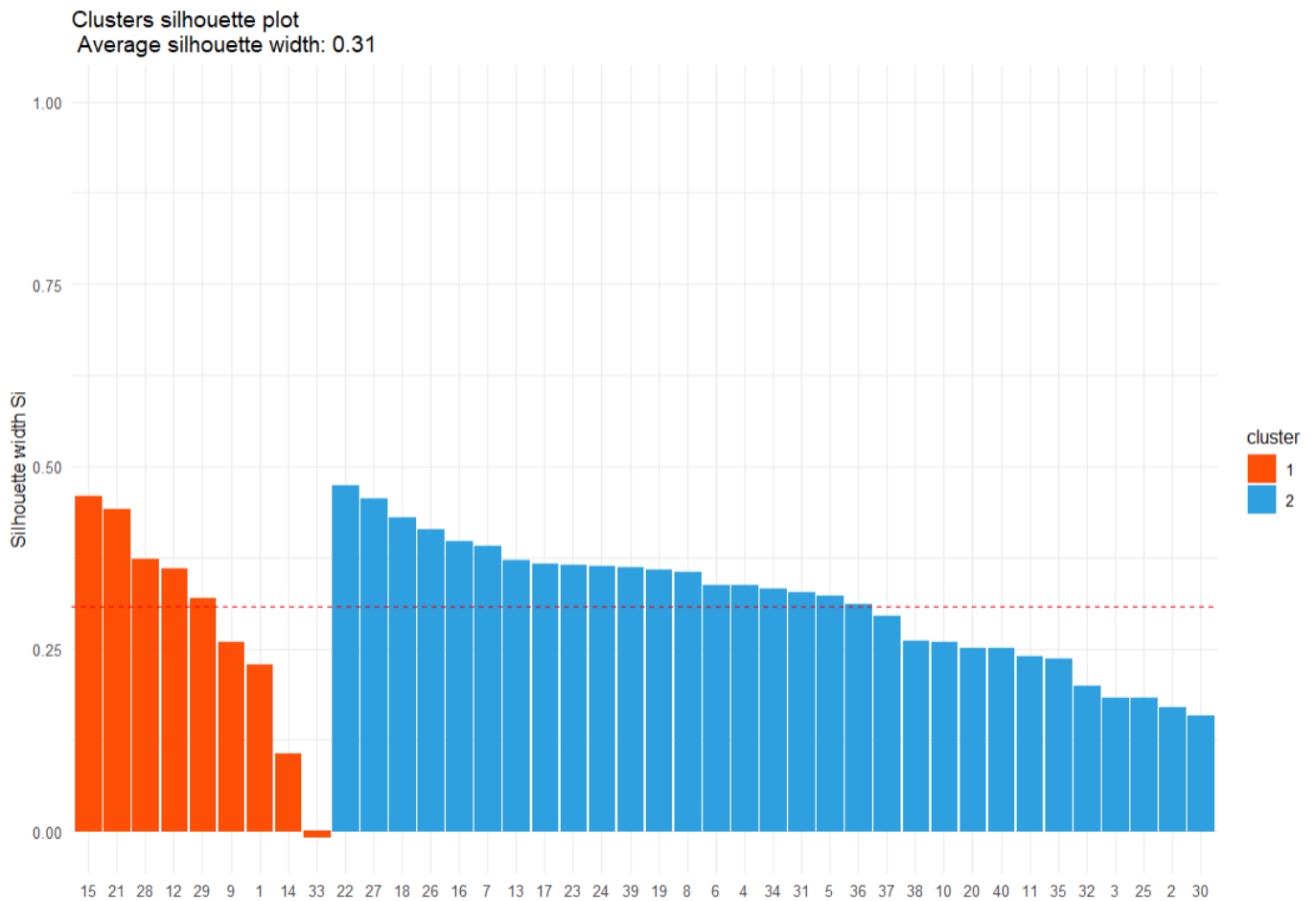*Figure 33: Clusters silhouette plot*

In the k-means algorithm the clustering provides an initial separation of relevant data into Hs_max -Q_max groups. The hierarchical clustering explains the relationship between data in the groups. This approach can be repeated with a different number of clusters to see if any improvement in the silhouette coefficient occurs (i.e., higher coefficient).

4. ANALYSIS OF WEATHER CIRCULATION TYPES

The objective of this activity is to analyse Circulation Types Classifications (CTCs) and the associated conditions for severe weather and related events, such as floods, storms and waves.

Atmospheric circulation refers to the large-scale movement of air within the atmosphere. By this, moisture and thermal energy are redistributed across Earth, as it moves from high pressure (anticyclones) areas to low pressure (depressions). It is responsible for daily changes in the weather. The persistence or recurrence of such weather patterns during seasons or months is crucial in the definition of anomalies in climate variables (ESOTC, 2022).

Weather or circulation type classifications (WTCs or CTCs) are widely used to analyze weather and climate conditions as a tool to separate the state of the atmosphere into subsets with similar and recurrent characteristics. Weather types are related to atmospheric characteristics closer to weather conditions such as instability and vorticity, while the so-called Circulation Types are directly linked to strictly circulation variables such as geopotential height and pressure.

Weather or circulation type classifications are also used to study the frequency and persistence of circulation types under various climate scenarios and assess the risk of severe weather in the coming decades.

During the 2000 to 2010 period, the relation between circulation weather types and extreme event have been increasingly explored (e.g., floods, Prudhomme & Genevier, 2011), owing to the impact of weather circulations on climate and particularly on storms.

Classifications of weather circulation have a long history in meteorology and climatology, starting with manual classifications (Hess & Brezowsky, 1952; Lamb H, 1972). In the last decades, the improvement of computing resources has fostered the development of fast and objective methods that used gridded datasets (e.g., of mean sea level pressure - MSLP or of geopotential height at 500 hectoPascal (500hgt)) to group the data into a small number of distinct categories.

In this context, the European project "Harmonization and Applications of Weather Types Classifications for European Regions" – COST Action733 (Fernández-Montes et al., 2014; Huth, 2010; Huth et al., 2008; Philipp et al., 2010) made an important contribution by providing a software code for classification of atmospheric circulation with different clustering methods. The choice of the best classification depends on several factors, such as the variable used for the classification (e.g., MSLP or 500hgt), the meteorological or environmental variable to be analyzed (e.g., temperature, precipitation, waves), the number of classes or clusters and the domain size.

LaMMA Consortium gained a considerable experience with this software, specifically about the choice of the most suitable classifications for describing the variability of surface temperature and precipitation in Italy (Iannuccilli et al., 2021; Vallorani et al., 2018). This experience, together with another preliminary study carried out on sea wave height data along the north Tyrrhenian and Ligurian sea, is very helpful in the identification of the most suitable classifications, in particular for the Italian domain, in order to evaluate the risk of severe weather or related phenomena (e.g., floods, storms, waves) by 2100, induced by changes in frequency and/or internal characteristics of circulation types, computed on the projections of several Regional Climate Models (RCMs) from the EURO-CORDEX dataset.

The methodology we have identified is being developed for the Italian peninsula, and exportable to other domains of equivalent size over Europe, to cover the target regions of the SCORE project. It is the result of previous analyses that allow to identify optimal classification solutions, among the (theoretically) several different possible ones. From previous research along the Italian peninsula, PCT09 (Principal Component analysis in Transversal-mode) and SAN09 (Simulated ANnealing), computed on mean sea level pressure (MSLP) were identified to be the best performing classifications on daily precipitation and maximum sea wave height respectively. They are both calibrated on ERA5 reanalysis datasets of MSLP over the reference period 1979 – 2005, using COST733 software package 1.4. The two CTCs are both made of 9 classes and were respectively computed with an obliquely rotated in PCT and with a k-means clustering SAN. Meteorological patterns of each Circulation Type (CT) are evaluated through monthly composite maps of rainfall amount, mean wind speed and temperature for the domain (in this case Italy) computed on the E-OBS regular grid dataset (0.1-degree spatial resolution) available on Copernicus C3S (Cornes et al., 2018) website (https://surfobs.climate.copernicus.eu/dataaccess/access_eobs.php). A punctuality analysis can also be carried out on long time series of observed data from weather stations.
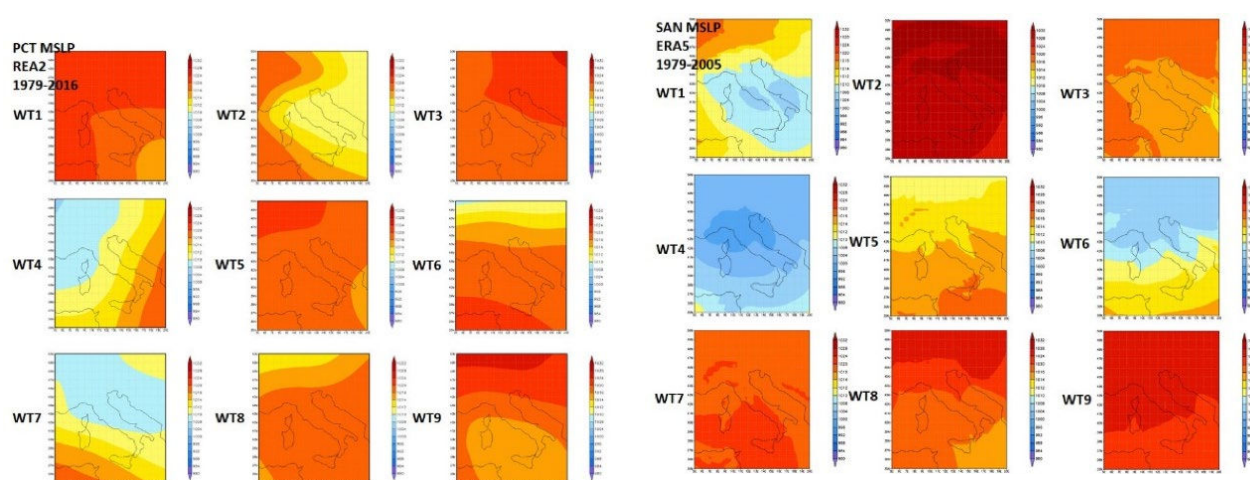

Figure 34: The two Circulation Type Classifications (CTCs) calibrated on ERA5 Mean Sea Level Pressure (MSLP) between the reference period 1979 – 2005; PCT09 is the CTC computed with Principal Component analysis in Transversal mode method (left panel) and SAN09 the CTC computed with k-means clustering Simulated ANnealing method (right panel).

A crucial part of the RCMs analysis is the comparison between the two calibrated CTCs on ERA5 and the assigned types for historical EURO CORDEX EUR-11 model (CNRM-CM5-ALADIN53 and multiple RCMs) on the reference period 1979-2005 in terms of frequency and persistence of circulation types, following the methodology described in the scientific report n.91 of MeteoSwiss entitled "Climate Change and Circulation Types on the Alpine Region" (Rohrer et al., 2017). This activity attempts to give an estimate of the quality of the EURO-CORDEX RCMs concerning their ability to simulate historic climate pattern through the frequency of the circulation types.

Afterwards, projections of MSLP field for a single (CNRM-CM5-ALADIN53) or multiple RCMs EURO CORDEX EUR-11 models will be converted, through an assignment algorithm of the COST733 software, into projected circulation types up to year 2100, according to SSP2-4.5 and SSP5-8.5 IPCC scenarios. This activity will allow us to show the pattern variations through the variation of the frequency of the types of circulation.

Finally, it will be possible to carry out a trend analysis of circulation type frequency and within-type changes, with a multiple scenario and a multi-model ensemble approach on a seasonal or monthly basis, in order to evaluate changes in the risk of extreme/severe weather events on a local scale.

REFERENCES

Acero, F. J., Gallego, M. C., García, J. A., Usoskin, I. G., & Vaquero, J. M. (2018). Extreme Value Theory Applied to the Millennial Sunspot Number Series. *The Astrophysical Journal*, *853*(1), 80. https://doi.org/10.3847/1538-4357/aaa406

Alboukadel Kassambara, F. M. (2017). *factoextra : Extract and Visualize the Results of Multivariate Data Analyses*. Cloud R Project. https://cloud.r-project.org/web/packages/factoextra/readme/README.html

Anderson, C. W., Leadbetter, M. R., Lindgren, G., & Rootzen, H. (1984). Extremes and Related Properties of Random Sequences and Processes. *Journal of the Royal Statistical Society. Series A (General)*, *147*(5). https://doi.org/10.2307/2981706

Bader, B., Yan, J., & Zhang, X. (2018). Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate. *Annals of Applied Statistics*, *12*(1). https://doi.org/10.1214/17-AOAS1092

Becker, R. A., Chambers, J. M., & Wilks, A. R. (1988). The new S language: a programming environment for data analysis and graphics, Wadsworth and Brooks. In *Cole*.

Benstock, D., & Cegla, F. (2017). Extreme value analysis (EVA) of inspection data and its uncertainties. *NDT and E International*, *87*. https://doi.org/10.1016/j.ndteint.2017.01.008

Boano-Danquah, J., Sigauke, C., & Kyei, K. A. (2020). Analysis of Extreme Peak Loads Using Point Processes: An Application Using South African Data. *IEEE Access*, *8*. https://doi.org/10.1109/ACCESS.2020.3015259

Bommier, E. (2014). *Peaks-Over-Threshold Modelling of Environmental Data*. Uppsala University U.U.D.M.

Coles, S. (2001). *Threshold Models BT - An Introduction to Statistical Modeling of Extreme Values* (S. Coles (ed.); pp. 74–91). Springer London. https://doi.org/10.1007/978-1-4471-3675-0_4

Fernández-Montes, S., Seubert, S., Rodrigo, F. S., Rasilla Álvarez, D. F., Hertig, E., Esteban, P., & Philipp, A. (2014). Circulation types and extreme precipitation days in the Iberian Peninsula in the transition seasons: Spatial links and temporal changes. *Atmospheric Research*, *138*. https://doi.org/10.1016/j.atmosres.2013.10.018

Ferreira, B. Y. A., & De Haan, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *Annals of Statistics*, *43*(1). https://doi.org/10.1214/14-AOS1280

Ferro, C. A. T. and S. J. (2005). *Extremes of Stationary Time Series* (J. . G. Y. . S. J. and T. J. (eds. . Beirlant (ed.); pp. 369–428). https://doi.org/10.1002/0470012382.ch10

Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, *24*(2). https://doi.org/10.1017/S0305004100015681

G'Sell, M. G., Wager, S., Chouldechova, A., & Tibshirani, R. (2016). Sequential selection procedures and false discovery rate control. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *78*(2), 423–444. http://www.jstor.org/stable/24775345

Gilleland, E., & Katz, R. W. (2016). extRemes 2.0: An Extreme Value Analysis Package in R. *Journal of Statistical Software*, *72*(8), 1–39. https://doi.org/10.18637/jss.v072.i08

Grover, R., & Vriens, M. (2006). The Handbook of Marketing Research. In *Journal of the Royal Statistical Society.*

Series A (Statistics in Society) (Vol. 156, Issue 3). SAGE Publications, Inc. https://doi.org/10.4135/9781412973380

Gumbel, E. J. (1958). *Statistics of the Extremes*. Columbia Press.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Applied Statistics*, *28*(1). https://doi.org/10.2307/2346830

Hess, P., & Brezowsky, H. (1952). *Katalog der Grosswetterlagen Europas. Ber Dt Wetterdienst in de US-zone*. Bad Kissingen Germany.

Huth, R. (2010). Synoptic-climatological applicability of circulation classifications from the COST733 collection: First results. *Physics and Chemistry of the Earth*, *35*(9–12). https://doi.org/10.1016/j.pce.2009.11.013

Huth, R., Beck, C., Philipp, A., Demuzere, M., Ustrnul, Z., Cahynová, M., Kyselý, J., & Tveito, O. E. (2008). Classifications of atmospheric circulation patterns: Recent advances and applications. *Annals of the New York Academy of Sciences*, *1146*. https://doi.org/10.1196/annals.1446.019

Iannuccilli, M., Bartolini, G., Betti, G., Crisci, A., Grifoni, D., Gozzini, B., Messeri, A., Morabito, M., Tei, C., Torrigiani Malaspina, T., Vallorani, R., & Messeri, G. (2021). Extreme precipitation events and their relationships with circulation types in Italy. In *International Journal of Climatology* (Vol. 41, Issue 10). https://doi.org/10.1002/joc.7109

Jane. (2020). *Package "MultiHazard" code: Multivariate statistical modelling*. GitHub. https://github.com/rjaneUCF/MultiHazard

Jane, R., Cadavid, L., Obeysekera, J., & Wahl, T. (2020). Multivariate statistical modelling of the drivers of compound flood events in south Florida. *Natural Hazards and Earth System Sciences*, *20*(10), 2681–2699. https://doi.org/10.5194/nhess-20-2681-2020

Janet E. Heffernan. (2022). *Package 'ismev': An Introduction to Statistical Modeling of Extreme Values*. Cran R Project. https://cran.r-project.org/web/packages/ismev/ismev.pdf

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*(3), 241–254. https://doi.org/10.1007/BF02289588

Karr, A. F. (2017). Point processes and their statistical inference, second edition, revised and expanded. In *Point Processes and their Statistical Inference, Second Edition, Revised and Expanded*. https://doi.org/10.1201/9780203743027

Kendall, M. G. (1957). Rank Correlation Methods. *Biometrika*, *44*(1/2). https://doi.org/10.2307/2333282

Kiran, K. G., & Srinivas, V. V. (2021). A Mahalanobis Distance-Based Automatic Threshold Selection Method for Peaks Over Threshold Model. *Water Resources Research*, *57*(1). https://doi.org/10.1029/2020WR027534

Lamb H. (1972). *British isles weather types and a register of the daily sequence of circulation patterns 1861-1971* (GEOPHYS MEM; GB; DA 1972; NO 116; PP 1-85; BIBL 2P1/2). file:///C:/Users/snalakurthi/Downloads/GeophysMem_116.pdf

Lin, G. F., Lin, H. Y., & Chou, Y. C. (2013). Development of a real-time regional-inundation forecasting model for the inundation warning system. *Journal of Hydroinformatics*, *15*(4). https://doi.org/10.2166/hydro.2013.202

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, *1*.

Mann, H. B. (1945). Nonparametric Tests Against Trend. *Econometrica*, *13*(3). https://doi.org/10.2307/1907187

Martin Maechler, P. R. (2022). *cluster: "Finding Groups in Data": Cluster Analysis Extended*. Cran R Project. https://cran.r-project.org/web/packages/cluster/index.html

Ossberger, J. (2022). *Package "tea": Threshold Estimation Approaches*. Cran R Package. https://cran.r-project.org/web/packages/tea/tea.pdf

Philipp, A., Bartholy, J., Beck, C., Erpicum, M., Esteben, P., Fettweis, X., Huth, R., Paul, J., Jourdain, S., Kreienkamp,

F., Krennert, T., Lykoudis, S., Micaelides, S., Pianko-Kluczynska, K., Post, P., Rasilla, D., Schiemann, D., Schiemann, R., Spekat, A., & Tymvios, F. (2010). *Cost733 Working Group2 - Implementation and development of weather types classification methods. – cost733 Final Workshop Classifications in atmospheric sciences and their applications, present state & future directions.*

Pickands, J. (1975). Statistical Inference using Extreme Order Statistics. *The Annals of Statistics*, *3*(1).

Prudhomme, C., & Genevier, M. (2011). Can atmospheric circulation be linked to flooding in Europe? *Hydrological Processes*, *25*(7). https://doi.org/10.1002/hyp.7879

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*(C). https://doi.org/10.1016/0377-0427(87)90125-7

Saeed Far, S., & Abd. Wahab, A. K. (2016). Evaluation of Peaks-Over-Threshold Method. *Ocean Science Discussions*, *July*.

Sinclair, C. D., Spurr, B. D., & Ahmad, M. I. (1990). Modified anderson darling test. *Communications in Statistics - Theory and Methods*, *19*(10), 3677–3686. https://doi.org/10.1080/03610929008830405

Smith, E. P. (2002). An Introduction to Statistical Modeling of Extreme Values. *Technometrics*, *44*(4), 397–397. https://doi.org/10.1198/tech.2002.s73

Solari, S., Egüen, M., Polo, M. J., & Losada, M. A. (2017). Peaks Over Threshold (POT): A methodology for automatic threshold estimation using goodness of fit p-value. *Water Resources Research*, *53*(4). https://doi.org/10.1002/2016WR019426

Stasinopoulos, M. (2007). Generalized Additive Models: An Introduction with R. by S. N. WOOD. *Biometrics*, *63*(4). https://doi.org/10.1111/j.1541-0420.2007.00905_3.x

Stephens, M. A. (1974). EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, *69*(347). https://doi.org/10.1080/01621459.1974.10480196

Stoyan, D. (2006). *Fundamentals of Point Process Statistics BT - Case Studies in Spatial Point Process Modeling* (A. Baddeley, P. Gregori, J. Mateu, R. Stoica, & D. Stoyan (eds.); pp. 3–22). Springer New York. https://doi.org/10.1007/0-387-31144-0_1

Thompson, P., Cai, Y., Reeve, D., & Stander, J. (2009). Automated threshold selection methods for extreme wave analysis. *Coastal Engineering*, *56*(10). https://doi.org/10.1016/j.coastaleng.2009.06.003

Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, *18*(4). https://doi.org/10.1007/BF02289263

Vallorani, R., Bartolini, G., Betti, G., Crisci, A., Gozzini, B., Grifoni, D., Iannuccilli, M., Messeri, A., Messeri, G., Morabito, M., & Maracchi, G. (2018). Circulation type classifications for temperature and precipitation stratification in Italy. *International Journal of Climatology*, *38*(2). https://doi.org/10.1002/joc.5219