Nat. Hazards Earth Syst. Sci., 26, 6157–6175, 2026
https://doi.org/10.5194/nhess-26-6157-2026
© Author(s) 2026. This work is distributed under
the Creative Commons Attribution 4.0 License.

Natural Hazards and Earth System Sciences
Open Access
EGU

Improved estimation of extreme sea levels via non-asymptotic statistical methods for coastal hazard assessment

S. Sithara[1], **Chiara Favaretto**[1], **Piero Ruol**[1], **and Marco Marani**[1,2]

[1]Department of Civil, Environmental, and Architectural Engineering, University of Padova, Padova, 35131, Italy
[2]Research Center on Climate Change Impacts, University of Padova, Rovigo, 45100, Italy

Correspondence: S. Sithara (sithara.sasidharan@unipd.it)

Received: 5 March 2026 – Discussion started: 16 April 2026
Revised: 7 October 2026 – Accepted: 5 November 2026 – Published: 25 November 2026

Abstract. Because sea level records are often short, non-asymptotic extreme value modeling is more accurate, as it does not depend on large-sample limiting assumptions. However, selecting independent events (IEs) is critical and depends on the chosen sea level threshold (SLT) and time window width (TWW) used. Existing literature often lacks a definitive methodology for IE selection. This study aimed to fill this gap by proposing two methodologies for selecting IEs (overlapping and sorting approaches) and a methodology to find the optimal combination of TWW and SLT. The identified IEs were employed to model extreme sea level events utilizing the Peak Over Threshold-Generalized Pareto Distribution (POT-GPD) and the metastatistical extreme value distribution (MEVD) approaches. A cross-validation approach, along with statistical metrics, was used to rigorously assess performance. Findings indicate that MEVD outperforms POT-GPD and is effective in estimating extreme events with longer return periods (RPs). Overall, MEVD with the overlapping method proved more effective at predicting long RP events, although MEVD with the sorting approach had comparatively lower uncertainty.

1 Introduction

Sea level is a critical coastal-hazard parameter that responds to a complex combination of interacting processes occurring in the atmosphere, in the ocean, and on the land surface (Ribeiro et al., 2014). In particular, extreme sea levels pose significant risks at vulnerable locations, such as low-lying coastal regions, leading to extensive damage to ecosystems as well as human communities and infrastructure (Tebaldi et al., 2021). Extreme sea levels are predominantly driven by the combination of tides and storm surges (meteorologically-induced increase in sea level), with wave setup further elevating water levels near shore (Kirezci et al., 2020). Recent studies reported that coastal flooding potentially impacts approximately 700 million people in low-elevation coastal zones, where around USD 13 trillion of global wealth is generated (Kirezci et al., 2023). With the ascending global temperatures, an acceleration in global mean sea-level rise is expected (Sithara et al., 2024), which, alone, has the potential to increase the frequency of coastal flooding, even in the absence of changes in waves and storm surges (Tebaldi et al., 2021). Additionally, changes in storminess may significantly influence the frequency and magnitude of sea level extremes (Woodworth et al., 2011; Caruso and Marani, 2022). The compound effect of these changes has indeed increased the risk of coastal flooding worldwide over the past few decades, a pattern projected to continue as a result of climate change (Seneviratne et al., 2021). In this context, reliable estimates of extreme sea levels are critical for informed decision-making and coastal protection (Ribeiro et al., 2014).

The widely adopted approach to estimate extreme events is the classical Extreme Value Theory (EVT). The classical EVT originated in the early twentieth century and describes the probability distribution of maxima and minima of event magnitudes within a time interval (block) of fixed duration (often taken to be one year). The classical EVT leads to a single, summarized analytical form known as the Generalized Extreme Value (GEV) distribution (Von Mises, 1936), which encompasses three possible limiting distributions (Fisher and Tippett, 1928; Gnedenko, 1943). This method either uses a limited number of block maxima (BM) (Coles, 2001; Smith,

2001; Hamdi et al., 2014) or a few exceedances of large values above a high threshold, termed peak over threshold (POT) (Pickands, 1975; Martins and Stedinger, 2001), to fit the GEV distribution. The BM-GEV and POT-GEV approaches thus both refer to an asymptotic GEV block-maximum distribution, but use different data to obtain distributional parameter estimates, i.e., just the yearly maxima or all exceedances over a high threshold, thus producing different numerical results. This traditional EVT is the basis for a large body of literature and applications in a wide range of disciplines, including storm surges (Hamdi et al., 2014; Vousdoukas et al., 2016), droughts (Sousa et al., 2011; Potisomporn et al., 2024), hurricanes (Coles and Casson, 1998), and many more. A significant challenge when using the traditional EVT-GEV methodology in investigating extreme sea levels is the scarcity of observations, both in terms of quantity and quality (Bernier et al., 2024; Fouad et al., 2025). Under these circumstances, the BM-GEV approach is critically inappropriate because it discards most of the already scarce available information and only makes use of block (typically yearly) maxima, resulting in an inefficient use of the data (Coles and Tawn, 2005; Haigh et al., 2010). The POT-GEV framework exploits information somewhat more effectively than the BM-GEV approach (Coles, 2001). Conveniently, when more than just a single yearly event is selected using a sufficiently large threshold, the distribution of the exceedances above the threshold converges to a Generalized Pareto Distribution (GPD) and is hence termed POT-GPD (Davison and Smith, 1990). However, estimation uncertainty is highly reliant on the threshold selection. A threshold that is too high will retain only a few peak values above the threshold, resulting in considerable error variance. On the other hand, a low threshold may select values that deviate significantly from a GPD distribution and may contradict the independence hypothesis of the framework (Coles, 2001). The use of a feasible intermediate threshold, e.g., in daily rainfall studies, typically leads to retaining a few events per year, thus not dramatically increasing the amount of information used in the estimation.

Additionally, it may be noted that the classical EVT (BM-GEV and POT-GPD) assumes the asymptotic GEV form, without the possibility to test for convergence. Hence, the asymptotic theory may be fitting the "wrong" distribution to the data at hand if the actual distribution is far from its limiting form. To address these limitations of the traditional EVT, Marani and Ignaccolo (2015) introduced the Metastatistical Extreme Value Distribution (MEVD), which does not assume an a-priori asymptotic distribution, and, instead, derives the distribution of extremes directly from the finite sample of the observed events.

Non-asymptotic methods, while presenting several advantages, require the definition of ordinary events, i.e., events in the observational record that may be considered to be realizations of independent random variables. This is required by the basic assumptions of both asymptotic and non-asymptotic methodologies and entails the screening of observational time series so as to isolate only the values in the record that are not generated by the same physical event (e.g., the same storm). Independent event separation has been studied for variables such as precipitation, river discharge, significant wave height, and winds (An and Pandey, 2005; Gao et al., 2020; Agilan et al., 2021; Pan et al., 2022). In those cases, a clear identification of periods in which no event is active is quite straightforward. The identification of independent storm surges within sea-level records is more complex, as the total sea level value includes both tidal and surge contributions. In practice, independence is ensured by defining events based on a threshold value and a minimum time separation. However, a definitive methodology to select the appropriate threshold and time window remains elusive.

Here we study the performance of non-asymptotic (MEVD) approaches, relative to asymptotic GEV results, while expanding the MEVD application by integrating two novel independent event (IE) selection methodologies. The analysis uses a cross-validation method to quantify the predictive estimation uncertainty associated with different non-asymptotic/asymptotic methods and different independent event separation methodologies. To achieve these objectives, we analyzed hourly sea level observations from 39 tidal gauge stations distributed along the Mediterranean and Southern European Atlantic coastlines. These stations encompass diverse tidal and meteorological regimes, providing a suitable dataset for evaluating extreme sea-level models. The datasets analyzed were either from various national agencies or the Global Extreme Sea Level Analysis (GESLA) dataset. The shortest time series in the dataset analyzed here is 15 years long (2010–2024, Alcudia), while the longest record is 118 years long (1905–2022, Trieste). Using these observations, we identified the optimal SLT and TWW values for each station and assessed the predictive uncertainty of the statistical models via a cross-validation procedure involving independent calibration and validation samples.

The paper is organized as follows: Sect. 2 describes the study area and data used. Section 3 outlines the methodology employed in this application, and the results are elucidated in Sect. 4. In Sect. 5, conclusions are drawn.

2 Study Area and Data

This study focused on coastal regions in the Mediterranean Sea and in the Atlantic Ocean in Southern Europe (Fig. 1). Hourly sea level observations from 29 tidal gauges along the coastline of the Mediterranean Sea and 10 tidal gauges along the coastline of Atlantic Ocean in Southern Europe were utilized for the analysis (Table 1 and Fig. 1).

The Mediterranean basin undergoes substantial anthropogenic pressure, with an approximate population of 160 million on its coastlines (UNEP/MAP, 2020; Toomey et al.,

Table 1. Details of tidal gauge stations.

Station number	Station Name	Data availability	Country	Source
1	Alexandria	1958–1976	EGY	Global Extreme Sea Level Analysis (GESLA)
2	Catania	1999–2024	ITA	Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA)
3	Livorno	1999–2024	ITA	ISPRA
4	Civitavecchia	1999–2024	ITA	ISPRA
5	Bari	1999–2024	ITA	ISPRA
6	Venice – Acqua Alta Oceanographic Station (AAOT)	1983–2023	ITA	Institute of Marine Sciences of the National Research Council, Venice, Italy
7	Porto Torres	2000–2024	ITA	ISPRA
8	Ancona	1999–2024	ITA	ISPRA
9	Ravenna	1999–2024	ITA	ISPRA
10	Palinuro	1999–2024	ITA	ISPRA
11	Otranto	2001–2024	ITA	ISPRA
12	Cagliari	1999–2024	ITA	ISPRA
13	Trieste	1905–2022	ITA	ISPRA
14	Imperia	1999–2024	ITA	ISPRA
15	Ortona	1999–2024	ITA	ISPRA
16	Toulon	1993–2024	FRA	Réseaux de Référence des Observations Marégraphiques (REFMAR)
17	Monaco Fontvieille	1999–2024	FRA	REFMAR
18	Ajaccio Aspretto	2002–2024	FRA	REFMAR
19	Marseille	1999–2024	FRA	REFMAR
20	Sete	2008–2024	FRA	REFMAR
21	Port Vendres	2008–2024	FRA	REFMAR
22	Bakar	1950–2021	HRV	Međugorac et al. (2026)
23	Valencia	1993–2024	ESP	Puertos del Estado
24	Barcelona	1993–2024	ESP	Puertos del Estado
25	Alicante	1957–2018	ESP	Puertos del Estado
26	Melilla	2008–2024	ESP	Puertos del Estado
27	Malaga	1993–2024	ESP	Puertos del Estado
28	Alcudia	2010–2024	ESP	Puertos del Estado
29	Ibiza	2003–2024	ESP	Puertos del Estado
30	Coruna	1992–2025	ESP	Puertos del Estado
31	Santander	1992–2025	ESP	Puertos del Estado
32	Vigo	1993–2025	ESP	Puertos del Estado

Table 1. Continued.

Station number	Station Name	Data availability	Country	Source
33	Dunkerque	1997–2025	FRA	REFMAR
34	Port Tudy	1977–2025	FRA	REFMAR
35	Roscoff	1979–2025	FRA	REFMAR
36	Saint Nazaire	1984–2025	FRA	REFMAR
37	Cascais	1971–2005	PRT	University of Hawaii Sea Level Center (UHSLC)
38	Ponta Delgada	1996–2018	PRT	UHSLC
39	Funchal	1986–2013	PRT	UHSLC

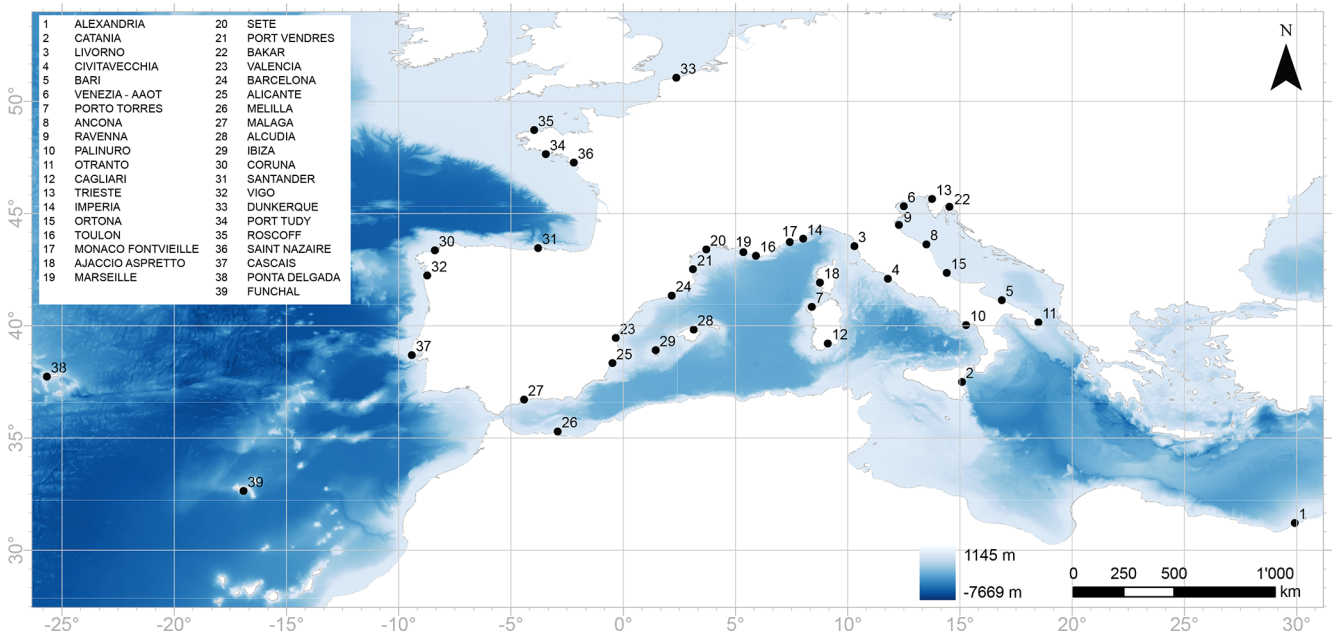


Figure 1. Tidal gauge locations (the bathymetry source: GEBCO 2025 Grid).

2025). Mediterranean coastlines are one of the most popular tourist destinations in the world. The irreplaceable cultural heritage sites attract one-third of international tourism flows. It also exhibits some of the world's busiest shipping routes (MedECC, 2020; Ali et al., 2022). Extensive infrastructures have been built in close proximity to the Mediterranean coasts due to the minimal tidal variation. The Mediterranean Sea is a global hot spot for climate change, with an above-average projected increase in temperatures (Toomey et al., 2025). Moreover, relative sea-level rise, which includes both changes in sea level and vertical land motion (e.g., land subsidence or uplift), varies considerably across the Mediterranean Sea. Satellite altimetry observations indicated a mean increase of $2.6 \pm 0.2\,\mathrm{mm\,yr^{-1}}$ in absolute (geocentric) sea level across the Mediterranean Sea over 1993–2015 (Ibáñez and Caiola, 2021). The Mediterranean basin experiences the highest rates of cyclogenesis globally. Hence, coasts are highly exposed to extreme sea level events, whose impacts are amplified by rising mean sea level (Toomey et al., 2022), exacerbating coastal hazards such as flooding, shoreline retreat, and saltwater intrusion, and posing a significant threat to heavily populated cities and important infrastructure. Given the high population density and economic significance of the Mediterranean coasts, it is essential to understand the likelihood of extreme sea level events.

Hundreds of thousands of people live along the Atlantic coastline in Southern Europe, which hosts several high-value economic activities and extensive infrastructures that must be adapted to intensifying natural coastal hazards, including marine flooding (Breilh et al., 2014; Antunes et al., 2019).

The southern European Atlantic coast is highly exposed to storm surges and large tidal ranges. Previous studies have shown that this coastline experiences some of the most extreme sea levels in southern Europe, driven by the combined influence of astronomical tides and intense meteorological forcing associated with Atlantic storm systems (e.g., Marcos et al., 2009). These factors lead to substantial spatial variability in extreme sea levels and frequent occurrences of coastal flooding events.

There is considerable variability in observed extreme sea levels, with values along the Atlantic coastline exceeding 250 cm above datum, while those in the Mediterranean are on average around 50–60 cm above the local datum, with high values of the order of 150–160 cm at numerous locations (e.g., Venice). Moving from the Bay of Biscay to Portugal and the southern coasts of Spain, the maximum observed values decrease by almost a meter. This spatial distribution is primarily driven by different tidal regimes (Tsimplis and Shaw, 2010). Higher water levels for a given return period (200–300 cm for the 50-year return level) are observed along the Atlantic coasts due to the larger tides (Marcos et al., 2009).

3 Methodology

3.1 Trend removal

Sea level includes contributions from a deterministic astronomical tidal component, a stochastic meteorological surge, and a long-term trend in mean sea level. The latter introduces a non-stationary component that must be removed prior to performing an extreme value analysis. In this study, the total sea level arising from the combined effects of astronomical and meteorological contributions was considered for all analyses, assuming the storm climate to be stationary over the period of record. From a statistical point of view, tides and surges could be analyzed separately. However, adding them back to obtain the total sea level requires assumptions regarding their independence, and the final result is affected by the accuracy with which the separation of these components could be obtained. Hence, attempts to separate the two signals are not without consequences for the subsequent analyses. In this study, we chose not to separate tidal and surge components, implicitly including tide–surge interactions and treating their sum as a single random variable. According to this approach, the observed total sea level $l(t)$ at any instant of time t can be represented as (Caruso and Marani, 2022):

$$l(t) = \mathrm{MSL}(t) + a(t) + m(t) \tag{1}$$

where $\mathrm{MSL}(t)$ is the mean sea-level measured with respect to an arbitrary datum, $a(t)$ is the astronomical tide and $m(t)$ is the storm surge, both measured with respect to the mean sea-level at time t.

$\mathrm{MSL}(t)$ changes because of long-term fluctuations in global sea levels, due to water volume changes, and to possible changes in datum elevation, e.g., if this is chosen to coincide with the soil surface, as a result of possible land uplift or subsidence. The tidal contribution, $a(t)$, caused by astronomical factors and by coastal morphology, contains periodic components with periods from 6 h to 18.61 years, the lunar nodal cycle (Valle-Levinson et al., 2021). The contribution of storm surge $m(t)$ is associated with the spatial distribution of atmospheric pressure and with wind direction and speed (Conte and Lionello, 2014; Bajo et al., 2023).

The sum of the tidal and surge components, subsequent to the subtraction of the mean sea-level constitutes the detrended sea level, $s(t)$:

$$s(t) = l(t) - \mathrm{MSL}(t) \tag{2}$$

Because sea level records at numerous tidal gauge stations were short, the detrended sea level here is obtained by simply removing a linear trend computed from the entire observed time series, as higher-order nonlinear trends cannot be reliably identified from such short records. After removing this long-term trend from sea-level observations, it is possible to study the detrended sea level as the sum of storm-surge and tidal contributions. The detrended sea level data were further processed to eliminate (i) years with less than 6 months of sea level observations, and (ii) measurement errors, i.e., the observations during periods that are not coherent with the regime of the remaining time series.

3.2 Extreme value theory

3.2.1 Asymptotic Extreme Value Theory

The standard extreme value theory (EVT) studies the distribution of maxima from data. Let us consider T independent and identically distributed (IID) random variables, X_1, X_2, …, X_T, with a common distribution, the "ordinary value" distribution $F(x)$ (here, an upper-case letter refers to the random variable, and a lower-case letter indicates the value of the random variable). We next divide the sample into non-overlapping blocks of equal size n and, for each block, we study $Y = \max(X_1, \ldots, X_n)$, the block maximum. The cumulative distribution function of block maxima can be expressed as, $P(Y \leq x) = F(x)^n$.

Under certain circumstances, there exist scaling constants $a_n > 0$ and b_n such that the normalized maximum converges to a nondegenerate distribution $H(y)$ (Smith, 2001). In other words, after rescaling, the maximum of a large sample can have a stable limiting distribution. The convergence occurs only when very large values become rarer at a specific rate as the sample size increases.

As the block size becomes infinitely large, the distribution of block maxima converges to one of three types of extreme value distributions according to the three-type theorem (Fisher and Tippett, 1928; Gnedenko, 1943): Gumbel,

Fréchet, or inverse Weibull distributions. These distributions can all be represented within a single framework using the generalized extreme value (GEV) distribution, as parameterized by Jenkinson (1955), and described in detail by Coles (2001).

$$H(y;\mu,\sigma,\xi) = \exp\left\{-\left[1+\xi\left(\frac{y-\mu}{\sigma}\right)\right]^{\frac{-1}{\xi}}\right\} \quad (3)$$

where, $\{y : 1+\xi(y-\mu)/\sigma > 0\}$, $\mu \in (-\infty,\infty)$ is the location parameter, $\sigma > 0$ is a scale parameter, and $\xi \in (-\infty,\infty)$ is the shape parameter. The case $\xi > 0$ represents a tail described by power-law (fat-tailed). The case $\xi = 0$ corresponds to an exponentially decreasing tail (light-tailed), while $\xi < 0$ is the case of a finite upper endpoint (short-tailed) (Batt et al., 2017; Smirlis et al., 2025). In practice, the GEV distribution is commonly fitted using the block maxima approach (BM-GEV). The BM-GEV approach is fitted to the yearly maxima extracted from the data series (Smith, 1989). An alternative to BM is the peaks over threshold (POT-GEV), focuses on exceedances over high thresholds instead of maxima over predetermined time periods (Smith, 1989; Caruso and Marani, 2022).

Suppose the exceedances, or differences between the observations over the threshold (u) and the threshold itself, are independent with a common distribution function F. For a sufficiently large threshold (upper endpoint), the distribution F of the exceedances $((Z-u)|Z > u)$ converges to the Generalized Pareto Distribution (GPD) given by (Davison and Smith, 1990)

$$G(y;\sigma,\xi) = 1-\left(1+\xi\frac{y}{\sigma}\right)^{\frac{-1}{\xi}} \quad (4)$$

where $\sigma > 0$ and ξ is arbitrary; the range of y is $0 < y < \infty$ if $\xi \le 0$, $0 < y < \frac{\sigma}{\xi}$ if $\xi > 0$. For sufficiently high u, there is some σ_u (depends on u) and some ξ for which the GPD is an excellent approximation to the exceedances distribution function F_u. Pickands (1975) already demonstrated this connection, and in that case, ξ in Eq. (4) is the same as that of the corresponding GEV distribution. Similar to GEV, there are three different cases based on the sign of ξ (Smith, 2001):

i. If $\xi > 0$, Eq. (4) is valid for $0 < z < \infty$ and the tail distribution function satisfies $1-G(y;\sigma,\xi) \sim cy^{-1/\xi}$ with $c > 0$, which is a traditional Pareto Tail.

ii. If $\xi < 0$, the G has an upper endpoint at $w_G = \sigma/|\xi|$, which corresponds to the inverse Weibull type of GEV.

iii. If $\xi = 0$, then $G(y;\sigma,0) = 1-\exp\left(-\frac{y}{\sigma}\right)$, which is the exponential distribution with mean σ leading to a Gumbel distribution in the GEV framework.

For a fixed u, if the number of exceedances is assumed to be a random variable, it can be modeled by leveraging the Poisson property of the exceedances, resulting in a model called the Poisson-GPD model. If the number (N) of exceedances above u in any one year has a Poisson distribution with mean λ, such that $N \ge 1$, the exceedances $Y_1, Y_2, \ldots, Y_N$ are IID from the GPD (Eq. 4). The probability that the annual maxima of the process less than a specified value z is:

$$Pr\left\{\max_{1\le i\le N} Y_i \le z\right\} = \exp\left\{-\lambda\left(1+\xi\frac{z-u}{\sigma}\right)^{-1/\xi}\right\} \quad (5)$$

Equation (5) is the same as Eq. (3), if $\sigma = \psi+\xi(u-\mu)$, and $\lambda = \left(1+\xi\frac{u-\mu}{\psi}\right)^{-1/\xi}$.

Here we used this approach, which is often termed POT-GPD, and, hereinafter, this terminology is used in this paper.

3.2.2 Non-asymptotic Extreme Value Theory

The MEVD approach developed by Marani and Ignaccolo (2015) relaxes the specific hypotheses under which the BM-GEV and POT-GEV approaches are derived and considers the parameters defining the cumulative distribution function of ordinary events as well as the number of ordinary events in a block as stochastic. Thus, the cumulative distribution of block maxima for the MEVD is defined as:

$$Q(x) = \sum_{n=1}^{\infty}\int_{\Omega_\Theta} F(x;\theta)^n q(n,\theta)\mathrm{d}\theta \quad (6)$$

where $q(n,\theta)$ is the joint probability distribution of event occurrences within a specified block and of the parameter vector, which is discrete in N and continuous in Θ; and Ω_Θ is the population encompassing all possible values of parameters.

For practical applications, the theoretical MEVD can be approximated by replacing the ensemble average in Eq. (6) with the sample average calculated over all the blocks in the time series. Furthermore, in this study, the MEVD was approximated by using the same parameters for the blocks while allowing n to vary in each block. Hence, the MEVD formulation adopted here can be expressed as

$$Q(x) \cong \frac{1}{M}\sum_{j=1}^{M} F(x;\theta)^{n_j} \quad (7)$$

where M is the number of blocks in the historical record and n_j is the number of events observed in the jth block. $F(x;\theta)$ is the cumulative distribution of ordinary event magnitudes, which is independent of a division in blocks: the entire calibration dataset here was used to estimate the parameters. As discussed by Zorzetto et al. (2016), in case one considers (i) x as the exceedance over a high threshold, (ii) $F(x;\theta_j)$ as GPD with fixed deterministic parameters, and (iii) n follows a Poisson distribution, then the GEV distribution can be derived as a particular case of the MEVD distribution.

Unlike the traditional asymptotic EVT, the MEVD approach is non-asymptotic and can leverage most of the available data to estimate the distribution parameters. The use of

a large proportion of observations afforded by the MEVD reduces the estimation uncertainty, particularly in the estimation of high quantiles, in comparison to the classical asymptotic EVT approaches (Caruso and Marani, 2022; Boumis et al., 2024). However, the MEVD formulation additionally requires prior understanding of the probability distribution that describes the underlying ordinary events.

The non-asymptotic MEVD adopted here uses a GPD for describing the ordinary event distribution, $F(x;\theta)$, following results from existing cross-validation analyses (Caruso and Marani, 2022). The ordinary-event GPD distribution parameters were estimated using the Probability Weighted Moments (PWMs) method (Greenwood et al., 1979). The PWMs approach is widely applied due to its reduced estimation bias and sensitivity to outliers in the data (Hosking, 1990).

3.3 Return Period

The primary purpose of the extreme value analysis is to determine the average recurrence interval or Return Period (RP) of specific event magnitudes. For a stationary process, the RP of a specific magnitude event is the average time between two consecutive exceedances of that magnitude. If the annual maximum magnitude s is surpassed, on average, once every RP years, the exceedance probability $E(s) = 1 - Q(s)$, in any given year is $E(s) = P[S \geq s] = 1/\mathrm{RP}(s)$. Consequently, the RP associated with a sea level value s is the reciprocal of the probability of exceedance and can be represented as a function of the cumulative distribution $Q(s)$ of yearly maxima. For the case of MEVD, RP can be expressed as: $\mathrm{RP}(s) = 1/E(s) = 1/1 - Q(s)$

For a specific value of MSL, there exists a one-to-one association between the aggregate of astronomical and storm surge contributions and the consequent total sea level ($l = s +$ MSL). Hence, it follows that $Q_s(s) = P[S > s] = P[S > l - \mathrm{MSL}] = P[L - \mathrm{MSL} > l - \mathrm{MSL}] = P[L > l] = Q_l(s)$. This implies that, once the cumulative distribution is known, the RP for the total sea level ($\mathrm{RP}(l)$) is:

$$\mathrm{RP}(l) = \frac{1}{1 - Q_l(s)} = \frac{1}{1 - Q_s(s)} = \frac{1}{1 - Q(l - \mathrm{MSL})} \quad (8)$$

3.4 Selection of independent sea level events

Independence of exceedances and the ordinary events is a basic requirement for the POT-GPD and MEVD approaches, respectively. The conventional approach used to ensure independence is a declustering process. In the case of POT-GPD, the procedure often involves choosing a sufficiently high threshold, and, if there are clusters of exceedances, it is treated as a single event with the maximum in the cluster as the representative. Threshold selection entails a trade-off between bias and variance; lower thresholds may introduce bias, while higher thresholds increase variance due to fewer instances of exceedance (Smith, 2001; Caruso and Marani, 2022). The original formulation of the MEVD, developed with application to daily rainfall, defined ordinary events based on a low threshold to isolate events generated by the same physical processes. Independence is in that context either an acceptable hypothesis for the original process (e.g., in the case of daily rainfall) or ensured using declustering approaches (Marra et al., 2018; Miniussi and Marani, 2020; Caruso and Marani, 2022).

In the present study, declustering was performed to identify independent events (IEs), which is based on site-specific characteristics, viz., sea level threshold (SLT) and time window width (TWW). The SLT choice depends on the magnitude of the local tidal range as well as storm contributions. The threshold is also selected to be sufficiently high to filter out water level peaks that are entirely influenced by tidal fluctuations, in the absence of any storm contribution (Caruso and Marani, 2022). Traditionally, the optimal choice of SLT is identified via goodness-of-fit measures (Davison and Smith, 1990) or mean excess plots (Smith, 2001), whereas a methodology for determining the optimal TWW remains elusive. Some studies concerning storm surge adopted a TWW between 24 and 72 h (Haigh et al., 2010; Bernardara et al., 2014; Cid et al., 2016).

The present study explored two declustering approaches for selecting IEs based on appropriate SLT and TWW values. While conceptually similar approaches have been used in POT frameworks (Solari et al., 2017; Martín et al., 2024), the use of threshold optimization methodologies to identify independent events in non-asymptotic extreme-value modeling methods, such as the MEVD, is here explored for the first time. First, all the down-crossings, i.e., intersections between the SLT and the detrended sea level line when the latter is characterized by a negative slope, are identified. Two subsequent down-crossing points define a temporal interval over which the hourly sea levels are above a specific threshold. The maximum sea level observed within this interval is defined as an event (blue points in Fig. 2a and b). The range of SLT values experimented with was set between the 90th percentile and 1.25 times the 99th percentile (to account for very high SLTs), further subdivided into equally spaced intervals, with a minimum width of 0.02 m.

Independence was then ensured by specifying an appropriate time window width (TWW) between successive events, yielding independent events (red points in Fig. 2a and b). The TWW values were adopted so that they spanned from 2 to 10 d across all tide gauges. A TWW of 6 d signifies 3 d prior to and 3 d after the event. The tidal regimes in the Mediterranean Sea and Southern European Atlantic Ocean are predominantly mixed semidiurnal and semidiurnal, respectively (Dias et al., 2021; Fanjul et al., 1998; Ferrarin et al., 2018). Consequently, the minimum TWW chosen was 2 d (across all stations), which is greater than the tidal cycle. An upper limit of 10 d was chosen, considering the range of the longest wind storm events (Besio et al., 2017; Amarouche et al., 2022; Borzì et al., 2024).

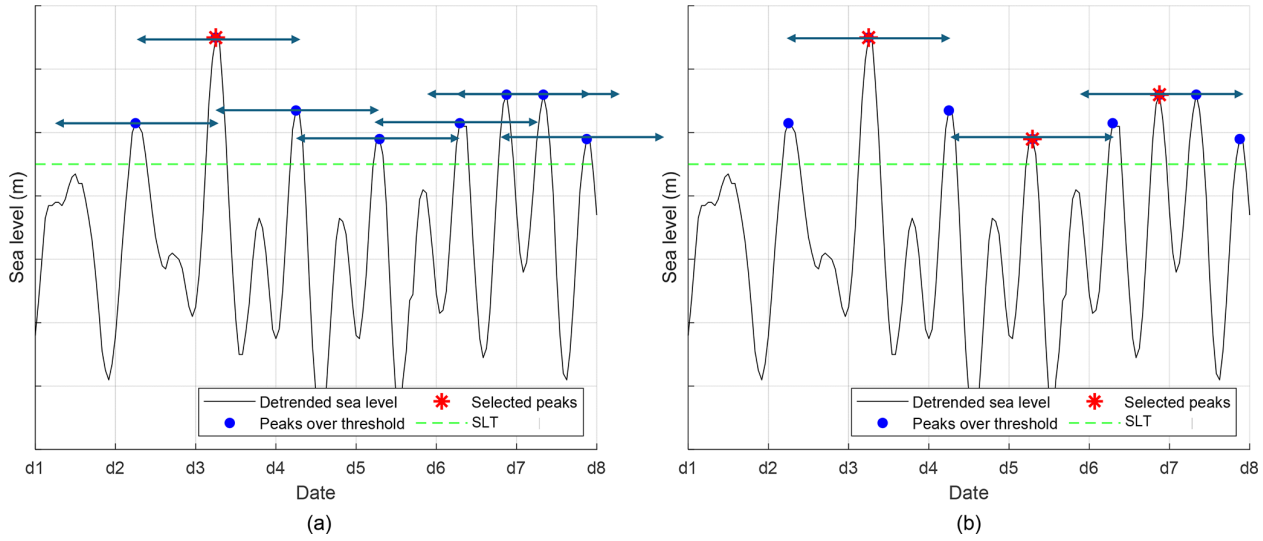


Figure 2. Selection of independent sea level events using **(a)** overlapping time window approach and **(b)** sorting approach. The green dashed lines represent the SLT, whereas the ↔ symbols depict the TWW.

After extracting the events based on the SLTs, two different methodologies were applied employing the TWWs to ensure independence: overlapping time window and sorting methods, as described below. Employing both methods, IEs were extracted for each SLT-TWW combination across all gauge stations. Here, the number of independent events selected for a specific combination of SLT and TWW is the same for both POT-GPD and MEVD: unlike in other studies and applications of non-asymptotic methods, differences on the results of the different extreme value methods do not depend on a difference in the amount of data that is utilized in the asymptotic vs the non-asymptotic approach. The differences only depend on the different formulations and assumptions underlying the two approaches.

The optimal SLT and TWW for each station for each extreme value model were identified via a cross-validation procedure, as explained in Sect. 3.5.1.

3.4.1 Overlapping time window method

The TWW is centered over each of the events, and wherever there is an overlap between the TWW of adjacent events, these events are merged into one (as seen in Fig. 2a). Among the merged events, the highest one (and the corresponding time) is extracted as representative (red asterisk in Fig. 2a). This process is carried out using an image processing algorithm named morphological dilation, utilizing a linear structuring element with a width proportional to the defined TWW.

3.4.2 Sorting method

The events extracted based on SLT are sorted in descending order of magnitude. Starting from the highest available event, all other events within the specified TWW from that event are excluded from the sorted list (Fig. 2b). For example, with a TWW of 6 d, once the first event in the sorted list is selected, events occurring 3 d prior to and 3 d following this event are eliminated. One then moves on to the next available event, the procedure was repeated, ultimately resulting in a set of IEs.

3.5 Optimal SLT and TWW selection and model performance evaluation through cross-validation

Accuracy assessment is critical to select and test a methodology. A common approach is to use goodness-of-fit metrics to compare the performance of various candidate distributions in describing observed samples (e.g., Chowdhury et al., 1991; Serinaldi and Kilsby, 2014; Krit et al., 2021). However, these goodness-of-fit metrics are not capable of indicating the accuracy with which the return period of an unseen extreme event can be estimated. In contrast, cross-validation approaches can provide out-of-sample estimates of uncertainty that more accurately characterize predictive uncertainty (Miniussi and Marani, 2020).

Here, we used a cross-validation procedure to (i) select the optimal SLT and TWW and (ii) evaluate the performance of the MEVD and POT-GPD approaches, particularly for high-quantile estimation. The cross-validation procedure involves a comparison of the probability of occurrence of the events predicted using the statistical models with the frequencies derived from observed data that are not used for model parameter estimation. This cross-validation was performed by dividing the available data into two subsets: calibration (sample used to estimate model parameters) and test data (used for comparison of model predictions with observations). As the key focus of this work is to confirm the reliability of the pro-

posed framework for yielding accurate results with short sea level records, we used a very short calibration sample size. Here, the calibration sample size was fixed at 4 years since it is the smallest RP that was able to compute for all the stations used in this study. The procedure can be summarized as follows:

i. The observational years in the record were randomly reshuffled while retaining all independent sea-level events within their original year of occurrence. This approach allows for the preservation of both the ordinary value frequency distribution for each year and the distribution of the number of events per year, while removing potential correlations within the time series.

ii. The observational sample was divided into two independent sub-samples by randomly selecting S years from the original record of length M, designated as the "calibration sample," which was utilized for parameter estimation. The remaining data, comprising $V = M - S$ years, is referred to as the "test sample," and was employed for testing.

iii. The empirical cumulative frequencies were computed using the Weibull plotting position formula $F_k = k/(V+1)$, where $k = 1, \ldots, M$ is the rank of the value x_k in an ascending order in a list of the annual maxima from the test sample. The return period (RP) associated with each yearly maximum is given by $\mathrm{RP}_k = 1/(1-F_k)$.

iv. The MEVD and GPD quantiles were estimated using the parameters obtained from the calibration sample.

v. The cross-validation scheme was repeated 1000 times to obtain a complete statistical description of the error metric.

3.5.1 Selection of optimal SLT and TWW

A value of the Root Mean Square Error (RMSE) was computed using the extreme sea level quantile estimates from the test sample for RP values less than or equal to the length of the calibration sample (here, 4 years), for all 1000 randomly generated cross-validation-test pairings. The optimal SLT-TWW combination to define the IEs was identified as the configuration that yielded the lowest RMSE. The evaluation of RMSE using shorter RP events was made because the study's primary goal is to assess the utility of the proposed methodology in data-scarce situations. In other words, whether reliable outputs can be achieved when the observational record is very short. This procedure allows us to determine whether the optimal SLT-TWW combination, chosen based on RMSE computed from quantiles with short RPs, can effectively be applied to estimate high quantiles. Here, for all the stations, the optimal combination of SLT and TWW was identified for POT-GPD and MEVD approaches integrated with two IE selection methodologies, yielding four distinct cases. These optimal SLTs and TWWs were used for the final modeling of extreme sea level events.

3.5.2 Model performance evaluation

The performance of the POT-GPD and MEVD methods was assessed graphically and via statistical metrics. Quantile–Quantile (Q–Q) plots were used to visually assess the agreement between observed and model-predicted extreme sea-level estimates. The slope of the best-fit line through the origin between the observed and model-predicted quantiles in the test sample (here, 1000 test samples; all plotting together) was used as a measure of agreement. The RMSE and Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970) were computed across all extreme sea level quantiles in the 1000 cross-validation test samples. NSE measures the predictive performance of a model by comparing the variance of prediction errors with the variability of the process as expressed by the variance of the observations. NSE values range from -∞ to 1, with 1 indicating perfect agreement, 0 indicating a performance that is equivalent to just using the observed mean as a prediction, and negative values indicating an inferior performance with respect to this reference. Additionally, a Quantile Ratio ($\mathrm{QR} = \frac{q_{\mathrm{predicted}}}{q_{\mathrm{observed}}}$) was calculated on the highest quantiles in the test samples to quantify the uncertainty in estimating the magnitude of longer RP events.

4 Results and Discussion

The independent sea level events identified for different combinations of thresholds and time window width employing the overlapping and sorting approach, which were subsequently used as input to the POT-GPD and MEVD approaches. The analysis encompasses four primary methodological approaches: (i) MEVD utilizing the overlapping time window method for the selection of independent events (MEVD-overlapping), (ii) MEVD employing the sorting method for independent event selection (MEVD-sorting), (iii) POT-GPD with the overlapping time window method (POT-GPD-overlapping), and (iv) POT-GPD utilizing the sorting method for independent event selection (POT-GPD-sorting). The results of the cross-validation procedure used to select the best SLT-TWW combination and to evaluate model performance are described below.

Among the 39 stations, results for four stations are presented here for detailed illustration: Venice-AAOT, Sete, Dunkerque, and Ponta Delgada. The four stations were selected as representative examples of the diverse tidal and storm regimes within the study domain. Together, these stations encompass a broad range of tidal amplitudes, storm climatologies, and sea-level variability: (i) Venice-AAOT is located in a shallow Adriatic Sea area dominated by storm surges from sirocco winds, low tidal range, but compound

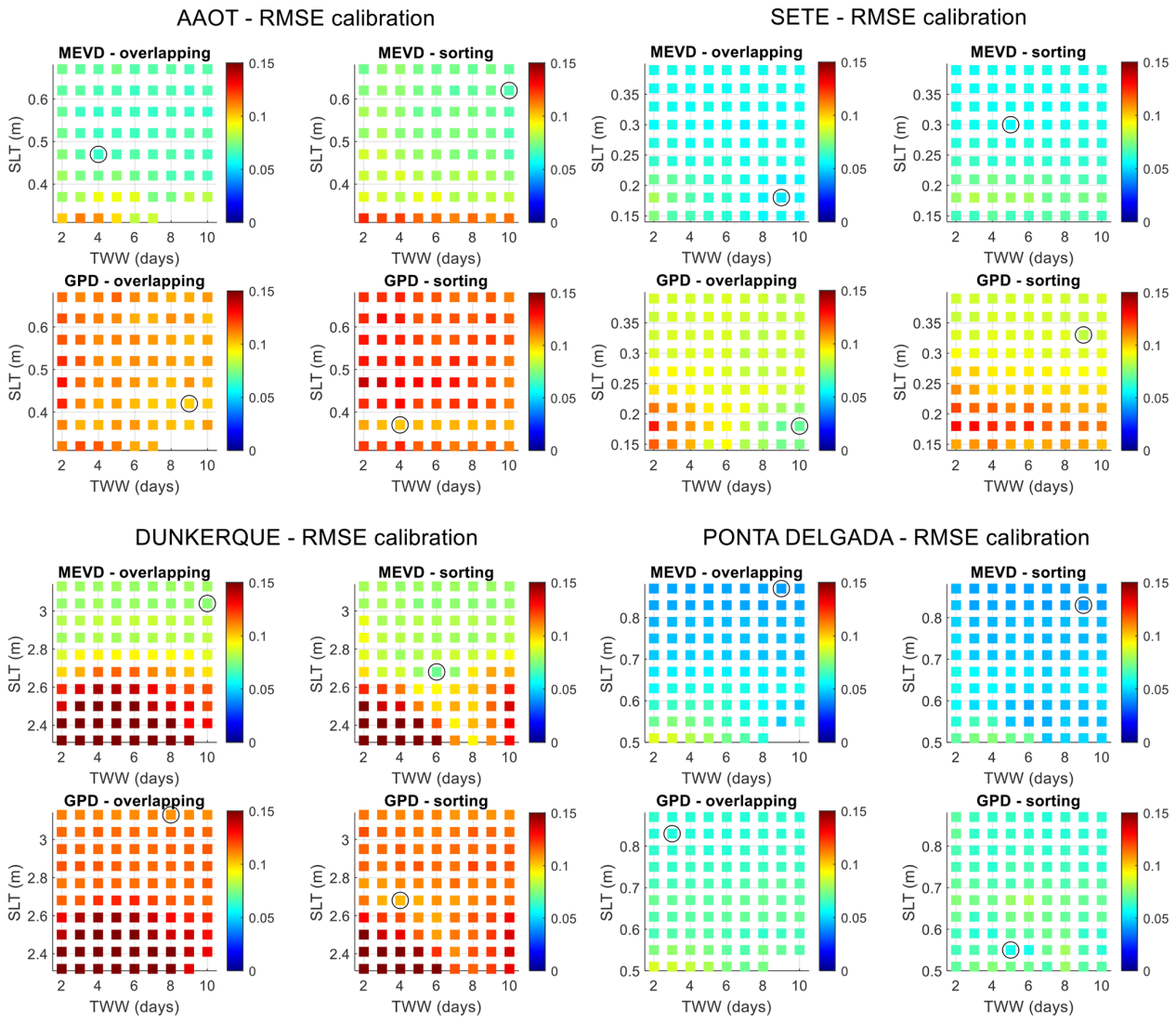


Figure 3. Heat maps indicating the best SLT-TWW combination based on the RMSE computed on all quantiles in the test sample with return periods less than or equal to 4 years during cross-validation.

tide-surge interactions, (ii) Sete is in the western Mediterranean sea, with a microtidal regime, affected by Mediterranean cyclones, (iii) Dunkerque is characterized by a large tidal range and a strong North Atlantic storm influence, (iv) Ponta Delgada is an open-ocean Atlantic island, exposed to North Atlantic swells and cyclones, with a relatively low tidal range. Figure 3 depicts the best SLT-TWW combinations selected for these stations, based on the cross-validation results. Extreme sea-level quantiles for RP less than or equal to the size of the calibration sample (here, 4 years) are estimated from calibrations on 1000 randomly selected training subsets. The deviations between the model-estimated quantiles and the corresponding observed quantiles of the test sample are then used to compute the RMSE values for each of the four methodological approaches being compared at each station. Notably, certain combinations were excluded from evaluation for the Venice-AAOT, Dunkerque, and Ponta Delgada stations due to the scarcity of annual maxima when employing the overlapping time window approach, as it was impossible to compute the maximum observed RP. As seen in the figures, for the Sete and Dunkerque stations, it is evident that the SLT and TWW selected for both the MEVD-overlapping and the POT-GPD-overlapping methodologies are quite close. Likewise, the optimal SLT and TWW values determined for the MEVD-sorting and POT-GPD-sorting approaches are similar. In contrast, the chosen SLT and TWW values for the other two stations (Venice-AAOT and Ponta Delgada) differ across methodologies. These findings indicate that the selected SLT and TWW values exhibit variability across different stations and methodologies, underscoring their dependence on local conditions. The longer TWW values identified at some stations may be associated with the

persistence of storm surges over several days and the occurrence of multiple exceedances associated with the same storm event. The selected TWW values may be analyzed in relation to the duration and clustering of meteorological and tidal forcings, as a lens to better understand the spatially varying drivers of storm surge extremes. An analysis that is left for subsequent developments and is not pursued further here. A sensitivity analysis was carried out by repeating the 1000 cross-validation analyses to obtain 20 independent realizations of extreme storm surge estimates for the four representative stations, to assess the sensitivity of the results to different optimal SLT–TWW combinations selected in different realizations. The results showed that the optimal SLT remained nearly unchanged across all methods except for the MEVD-sorting approach, where the largest difference (approximately 35 cm) was observed at Dunkerque, representing about 6 % of the local spring tidal range (~ 5.5 m). The optimal TWW exhibited variability across the 20 independent realizations for the four representative stations, with coefficients of variation (CVs) calculated from the selected TWW values ranging from 0 to approximately 60 %, depending on the station and approach. Furthermore, the variability in estimated extreme sea-level quantiles for a 100-year return period was assessed by computing the CV from the 20 quantile estimates obtained from individual realizations. The results revealed low variability (CV ≤ 5.47 %) for all four estimation approaches, indicating robustness with respect to the optimal SLT-TWW selection. This study found that in some cases MEVD approaches selected a higher SLT than those of POT-GPD approaches. Furthermore, it was interesting to note in the study carried out by (Caruso and Marani, 2022) that the SLT values used for the MEVD approach were not markedly smaller than those used for the POT-GPD method. In fact, the difference between the thresholds was $\sim$ 6–10 cm. Therefore, choosing a lower SLT and TWW may not invariably define the ordinary events appropriately.

The performance of the models with regard to extreme sea level was assessed graphically and using statistical metrics. Figure 4 shows the Q–Q plots of the extreme sea level events based on the optimal SLT-TWW selection for all methods for Venice-AAOT, Sete, Dunkerque, and Ponta Delgada stations, as illustrative and representative examples. Quantiles predicted by the MEVD approach match observed ones well, as evidenced by a good alignment along the 1 : 1 line in the Q–Q plots. This alignment appears tighter for MEVD predictions than for POT-GPD-predicted quantiles, both for high and low values of extreme surges (Fig. 4C and D). Similar results were observed for other stations. Additionally, for 33 out of 39 stations, the slope of the least-squares regression line through the origin is closer to 1 for the MEVD methods than for the POT approaches.

Further analyses were conducted to better understand the variability in SLTs and TWWs. Figure 5 illustrates the violin plots displaying the variation of the Quantile Ratio (QR) evaluated for the highest quantile in cross-validation experiments, across different methods and for various SLT values, while maintaining a fixed TWW value of 4, 9, 6, and 3 d for the representative stations, respectively. QR values vary significantly as a function of the thresholds, underscoring the role of threshold selection in extreme value analysis. QR values for the highest quantile show some variability as a function of TWW (Fig. 6). At some sites, e.g., Dunkerque, there is more noticeable variation in QR across the methods. However, the relatively consistent median QR values across different TWW values suggest that model predictions are less influenced by TWW compared to SLT (Fig. 6). Therefore, although longer TWW values may potentially combine a few closely spaced events, this does not significantly influence the estimated extreme sea level quantiles.

The performance of the four methods was also evaluated based on the QR computed on the highest quantiles as well as RMSE and NSE computed on all quantiles for all the stations. All the metrics were evaluated using the results obtained for the optimal SLT-TWW combination across stations. As seen in Figs. 7 and 8, the MEVD approaches outperform POT-GPD approaches across all stations in terms of RMSE and NSE. No negative values of NSE, which correspond to highly erroneous predictions, were found for MEVD. Analogously, there is a good agreement between the observations and MEVD predictions for the highest quantile (extreme sea level events associated with long RPs) across the stations, as shown in Fig. 9 (left). Figure 9 (right) illustrates the best method identified for each station, in terms of QR computed from high quantiles, against the longest RP evaluated at each station. From this figure, it is apparent that the MEVD approach can estimate longer return period events compared to the POT-GPD approach. From the heat maps (Fig. 3), it was already seen that MEVD methods used higher SLT compared to POT-GPD methods in certain instances. Therefore, it is clear that MEVD worked well also with a smaller number of events compared to POT-GPD methods. This could be attributed to the formulation of MEVD, which explicitly incorporates the number of events (n_j) as represented in Eq. (7).

The analysis was further extended to explore the role of the number of events per year utilized in fitting the MEVD and POT-GPD methods. Figure 10 compares the number of events per year used by MEVD and that used by the POT-GPD, along with performance information. MEVD outperforms POT-GPD at 25 out of 39 stations for the overlapping method and 33 out of 39 with the sorting method. Interestingly, as shown in Fig. 10 (left), the MEVD-overlapping approach consistently employed a number of events between $\sim$ 5–12 per year, regardless of the location. The MEVD-sorting approach used a smaller number of events ($\sim$ 5–20) per year compared to POT-GPD-sorting for all the stations except for 8 stations along the Atlantic coastline. For those 8 stations, the number of events used by MEVD-sorting aligns with POT-GPD-sorting. This observation suggests that the number of events required for the MEVD-

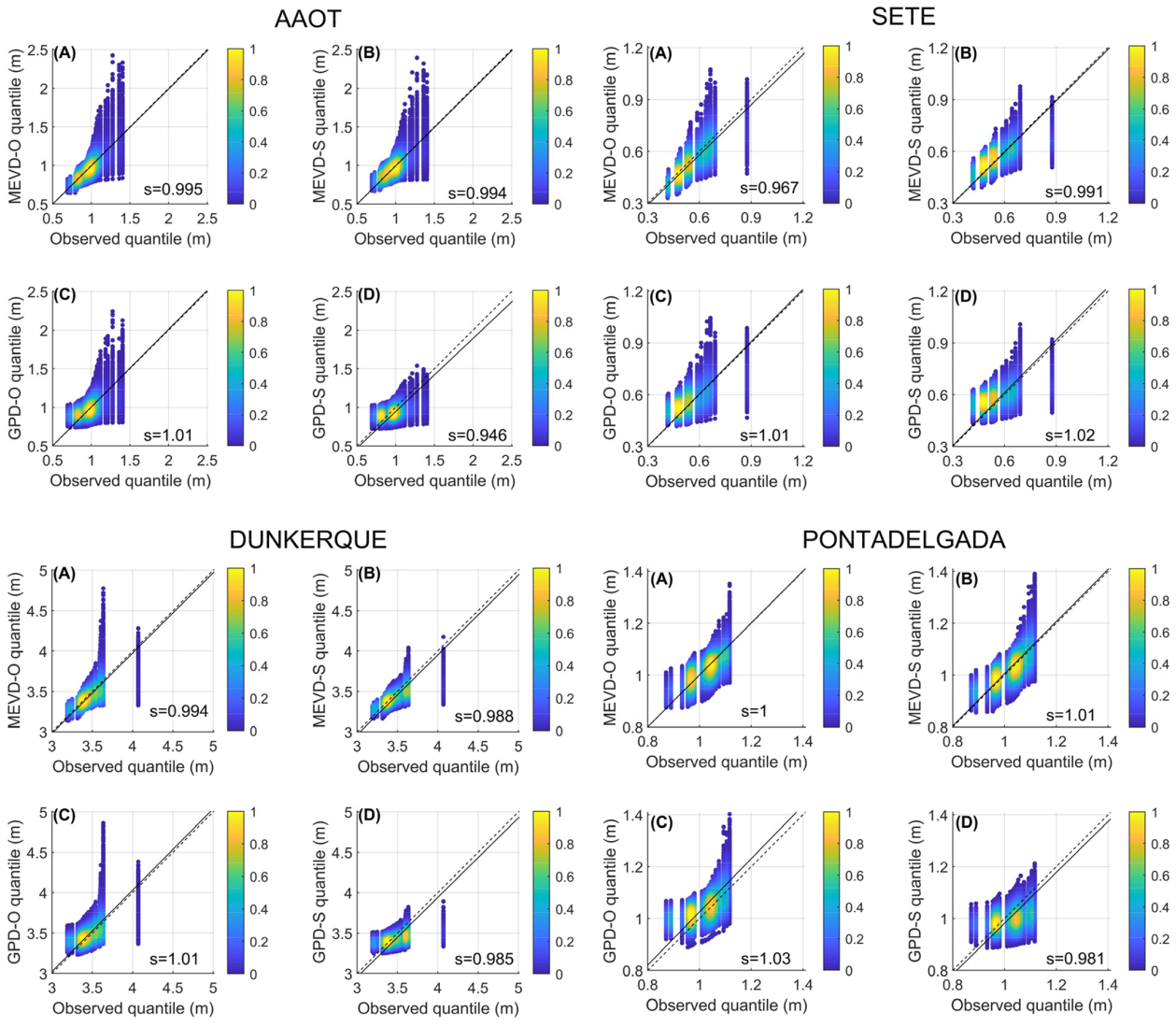


Figure 4. Q–Q plots of all sea level quantiles computed for the best combination of SLT and TWWs selected for each method. The plot shows the slope (s) of the best-fit line (solid black line) of the quantiles and the 1 : 1 line (dashed black line).

sorting approach is more site-specific. This can be attributed to local hydrodynamic processes or geometry, specifically the tidal dominance in the Atlantic Ocean, as well as the fact that the location of some of these stations near harbors results in moderate events. The sorting approach identifies more events than the overlapping approach, because the latter merges more events into one. Consequently, the MEVD-sorting approach uses a larger number of events per year than the MEVD-overlapping method to capture the process accurately. Focusing on SLT-TWW combinations that produce $\sim$5–12 events per year is advisable for analysis using the MEVD-overlapping approach. However, the optimal combination should be determined based on RMSE.

The estimation error (difference between predicted and observed values) associated with the MEVD-overlapping and MEVD-sorting approaches regarding high quantile estimation was evaluated across stations, as illustrated in Fig. 11. In the boxplot, the interquartile range is a measure of uncertainty. It is apparent from the figure that for both methods, the median of the estimation error is around zero. However, the prediction uncertainty is comparatively less for the MEVD-sorting approach. The differences in estimation-error variability across the Atlantic–Mediterranean study area likely reflect regional variations in the physical processes governing extreme sea levels. Along the Southern European Atlantic coast, extreme sea levels are influenced by the interaction of semidiurnal tides with energetic storm surges and waves associated with North Atlantic storms. In contrast, much of the Mediterranean Sea is microtidal and semi-enclosed, with extreme sea levels primarily driven by meteorological forcing, except in regions such as the Northern Adriatic, where tidal and storm-surge influences are more

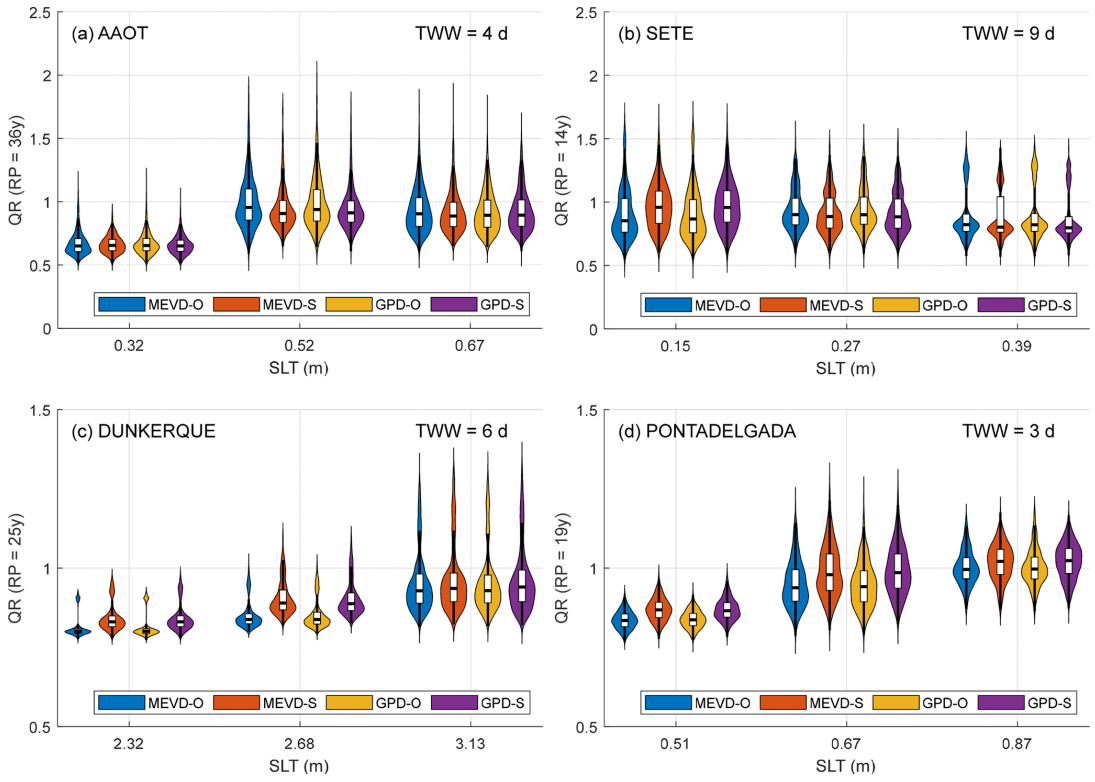


Figure 5. Violin plots showing variation of QR computed on the highest quantiles during the cross-validation for a fixed TWW for the stations **(a)** Venice-AAOT, **(b)** Sete, **(c)** Dunkerque, **(d)** Ponta Delgada.

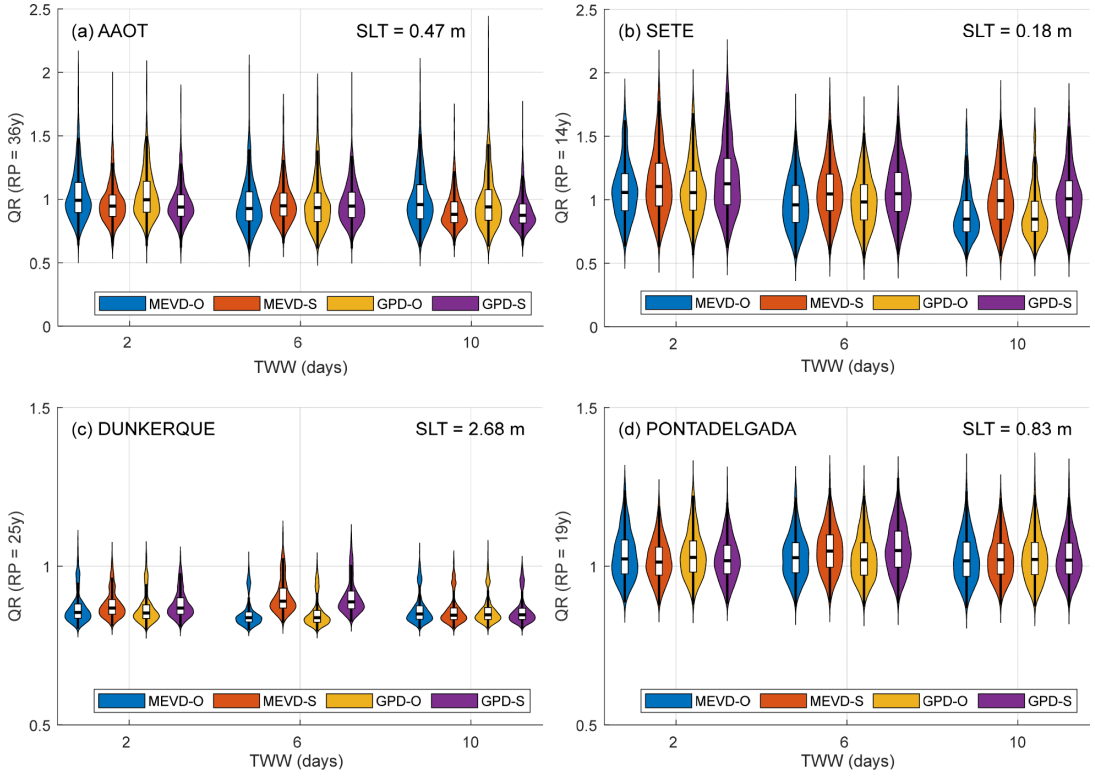


Figure 6. Violin plots showing variation of QR computed on the highest quantiles during the cross-validation for a fixed SLT for the stations **(a)** Venice-AAOT, **(b)** Sete, **(c)** Dunkerque, **(d)** Ponta Delgada.

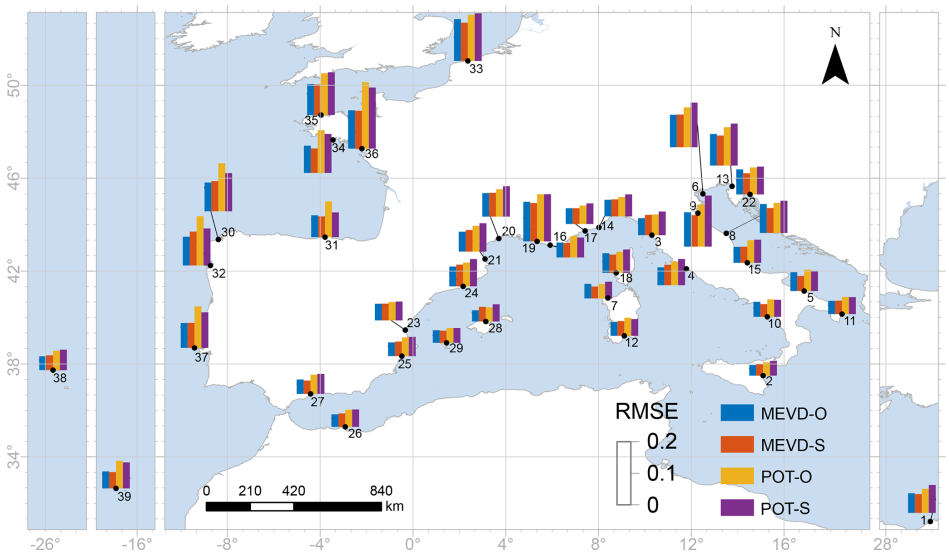


Figure 7. RMSE computed on all quantiles estimated for the best combination of SLT and TWW during cross-validation across all methods and stations.

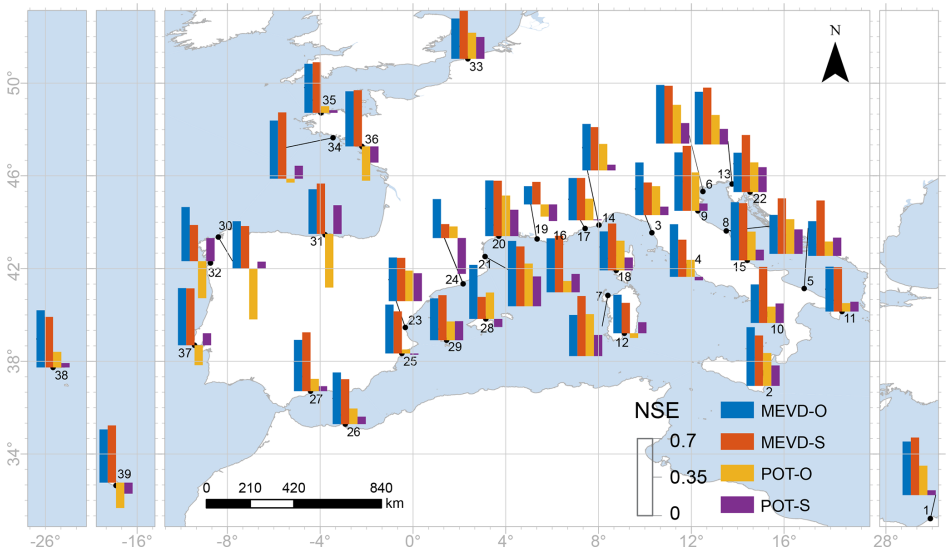


Figure 8. NSE computed on all quantiles estimated for the best combination of SLT and TWW during cross-validation across all methods and stations.

pronounced. The observed estimation-error variability, however, does not show a clear basin-scale pattern, indicating that it is likely controlled by a combination of local tidal, meteorological, and coastal characteristics rather than by basin-scale differences alone.

5 Conclusions

The comparative assessment MEVD and POT-GPD models, when integrated with novel independent event selection methodologies and applied to sea level observations across 39 stations along the Mediterranean and southern European Atlantic coastlines, provided insights into the predictive performance of the models. Our investigations quantitatively confirm the advantages of the MEVD over the conventional POT-GPD stemming from the conceptual differences between the methods. The quantitative evaluation of the predictive uncertainty of the statistical models with respect to high-quantile estimation was made possible by the systematic use of a cross-validation process. Selection of the optimal SLT and TWW based on RMSE values evaluated across all quantiles with RP $\leq$ calibration sample size (4 years), using the cross-validation procedure, revealed that selecting the appropriate SLT and TWW is very important for

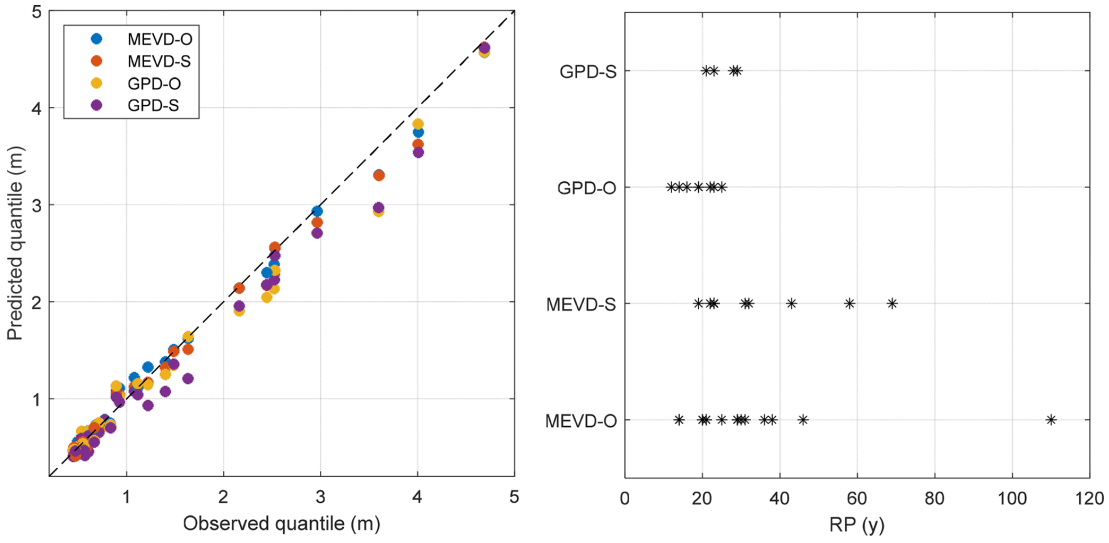


Figure 9. Observed vs. predicted highest quantiles across stations (left) and outperformed method (i.e. the one which yielded the best QR among the four) vs. maximum RP computed for each station (right).

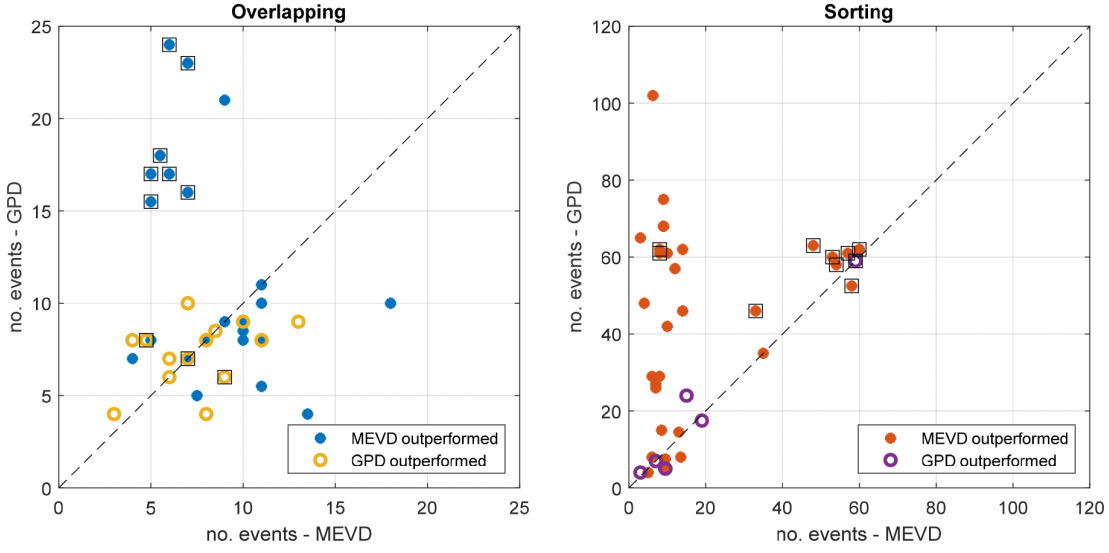


Figure 10. Number of events per year used by POT-GPD approach vs. those used by MEVD approach. The points enclosed in a square are the ones corresponding to the stations along the Atlantic coast.

defining independent ordinary events in the case of MEVD. However, the model's predictions are more sensitive to SLT choice in comparison to TWW. The optimal SLT-TWW combination selected based on the analyses performed using a very short record of observations could reliably model the high quantiles, as evidenced by the cross-validation results. A lower SLT and a minimum TWW will not invariably yield independent ordinary events within a block. With a fixed calibration sample size of 4 years, the comparative evaluation of four model configurations (a) MEVD-overlapping, (b) MEVD-sorting, (c) POT-GPD-overlapping, and (d) POT-GPD-sorting (IEs defined based on the optimal SLT-TWW combination) evaluated based on the statistical metrics (QR, RMSE, and NSE), indicates that MEVD approaches outperform POT-GPD approaches in terms of prediction accuracy in the context of data scarcity. We found that the MEVD approaches yield reliable estimates of high quantiles for almost all the tidal gauge stations examined, even with a small calibration sample size (4 years). Results indicated that MEVD approaches are more effective at estimating long RP events compared to POT-GPD approaches. Both POT-GPD approaches reliably estimated only extreme sea level events with RP less than 40 years, whereas the MEVD approaches predicted events with RP greater than 65 years. Specifically,

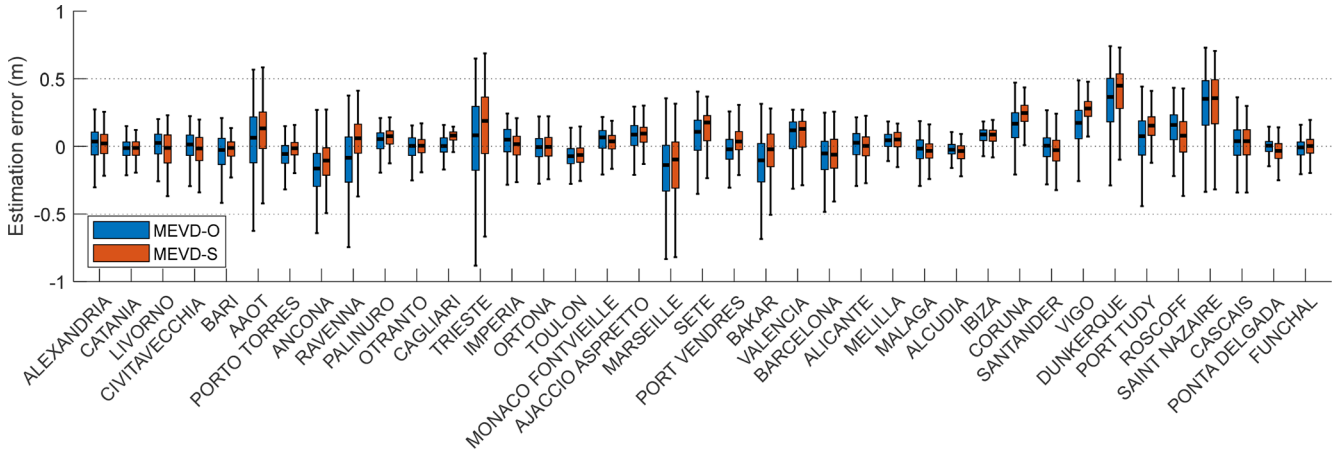


Figure 11. Boxplot of estimation error associated with the highest quantiles predicted using the MEVD-sorting method vs. those predicted using the MEVD-overlapping method.

the MEVD-overlapping method estimated events with RP greater than 100 years. A notable finding of the study is that the MEVD-overlapping approach consistently utilized a number of events per year between ∼ 5–12, regardless of the location. The MEVD and POT-GPD approaches integrated with the sorting method for IE selection utilized a smaller number of events per year, except for a few stations along the Atlantic coastline. This disparity could be linked to local hydrodynamic processes (tidal dominance) or specific geometric factors (e.g. in the presence of harbors). Furthermore, among the MEVD approaches, prediction uncertainty was comparatively less for the MEVD-sorting approach. Overall, it can be concluded that while the optimal IE selection methodology is slightly site-specific, the MEVD formulation facilitates estimation of longer RP events even in data-scarce contexts. The findings of this study help quantify the risk associated with rare extremes and design robust defense infrastructures. For practical applications for the Mediterranean coastline, the MEVD-overlapping approach is suggested with a TWW of 6–10 d and a SLT of 0.35 m in the Alboran Sea, 0.2–0.25 m in the Western basin, including the Tyrrhenian Sea, and 0.2–0.3 m in the Southern Adriatic Sea, up to 0.7 m in the Northern Adriatic Sea.

Code and data availability. All datasets used in the present study are publicly available. The sea level observations for the Italian stations, except for the AAOT tide gauge, were sourced from Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA) (https://www.mareografico.it/, last access: 16 November 2026) and for the AAOT from the Institute of Marine Sciences of the National Research Council (https://www.comune.venezia.it/it/content/3-piattaforma-ismar-cnr, last access: 16 November 2026). Data for the French stations were obtained from Réseaux de Référence des Observations Marégraphiques (REFMAR) (https://refmar.shom.fr/donnees-refmar-sur-data.shom.fr/telechargement-des-donnees, last access: 16 November 2026), while Spanish stations' data from Puertos del Estado (https://portus.puertos.es/#/?locale=en, last access: 16 November 2026). Egyptian and Portuguese stations' data were acquired from the University of Hawaii Sea Level Center (UHSLC) (https://gesla787883612.wordpress.com/downloads/, last access: 16 November 2026). Sea level observations of the Croatian station were retrieved from an open data repository SEANOE (DOI: https://doi.org/10.17882/85171, Međugorac et al., 2026). The code developed for the analyses is available from the authors upon reasonable request.

Author contributions. SS: Conceptualization, Methodology, Formal analysis, Software, Validation, Investigation, Data curation, Visualization, Writing (original draft preparation), Writing (review and editing); CF: Conceptualization, Methodology, Data curation, Validation, Visualization, Supervision, Writing (review and editing); PR: Conceptualization, Methodology, Validation, Supervision, Writing (review and editing); MM: Conceptualization, Methodology, Validation, Supervision, Funding acquisition, Project administration, Writing (review and editing).

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements. This study is carried out within the project "RETURN – Multi-risk science for resilient communities under a

changing climate", funded by the European Union – Next Generation EU, under the National Recovery and Resilience Plan (PNRR), Mission 4 "Education and research", Component 2 "From research to business", Investment 1.3 (identification code PE0000005, Spoke 8 – DS – Science underpinning Climate services for risk mitigation and adaptation, CUP code C93C22005160002).

Financial support. This research has been supported by the European Commission, NextGenerationEU (grant no. PE0000005).

Review statement. This paper was edited by Piero Lionello and reviewed by two anonymous referees.

References

Agilan, V., Umamahesh, N. V., and Mujumdar, P. P.: Influence of threshold selection in modeling peaks over threshold based nonstationary extreme rainfall series, J. Hydrol., 593, https://doi.org/10.1016/j.jhydrol.2020.125625, 2021.

Ali, E., Cramer, W., Carnicer, J., Georgopoulou, E., Hilmi, N. J. M., Le Cozannet, G., and Lionello, P.: Cross-Chapter Paper 4: Mediterranean Region, in: Climate Change 2022 – Impacts, Adaptation and Vulnerability, edited by: Tignor, M., Poloczanska, E. S., Mintenbeck, K., Alegría, A., Craig, M., Langsdorf, S., Löschke, S., Möller, V., Okem, A., and Rama, B., Cambridge University Press, Cambridge, UK and New York, NY, USA, 2233–2272, https://doi.org/10.1017/9781009325844.021, 2022.

Amarouche, K., Akpınar, A., and Semedo, A.: Wave storm events in the Western Mediterranean Sea over four decades, Ocean Model., 170, https://doi.org/10.1016/j.ocemod.2021.101933, 2022.

An, Y. and Pandey, M. D.: A comparison of methods of extreme wind speed estimation, J. Wind Eng. Ind. Aerod., 93, 535–545, https://doi.org/10.1016/j.jweia.2005.05.003, 2005.

Antunes, C., Rocha, C., and Catita, C.: Coastal flood assessment due to sea level rise and extreme storm events: A case study of the atlantic coast of Portugal's mainland, Geosciences, 9, https://doi.org/10.3390/geosciences9050239, 2019.

Bajo, M., Ferrarin, C., Umgiesser, G., Bonometto, A., and Coraci, E.: Modelling the barotropic sea level in the Mediterranean Sea using data assimilation, Ocean Sci., 19, 559–579, https://doi.org/10.5194/os-19-559-2023, 2023.

Batt, R. D., Carpenter, S. R., and Ives, A. R.: Extreme events in lake ecosystem time series, Limnol. Oceanogr. Lett., 2, https://doi.org/10.1002/lol2.10037, 2017.

Bernardara, P., Mazas, F., Kergadallan, X., and Hamm, L.: A two-step framework for over-threshold modelling of environmental extremes, Nat. Hazards Earth Syst. Sci., 14, 635–647, https://doi.org/10.5194/nhess-14-635-2014, 2014.

Bernier, N. B., Hemer, M., Mori, N., Appendini, C. M., Breivik, O., de Camargo, R., Casas-Prat, M., Duong, T. M., Haigh, I. D., Howard, T., Hernaman, V., Huizy, O., Irish, J. L., Kirezci, E., Kohno, N., Lee, J. W., McInnes, K. L., Meyer, E. M. I., Marcos, M., Marsooli, R., Martin Oliva, A., Menendez, M., Moghimi, S., Muis, S., Polton, J. A., Pringle, W. J., Ranasinghe, R., Saillour, T., Smith, G., Tadesse, M. G., Swail, V., Tomoya, S., Voukouvalas, E., Wahl, T., Wang, P., Weisse, R., Westerink, J. J., Young, I., and Zhang, Y. J.: Storm surges and extreme sea levels: Review, establishment of model intercomparison and coordination of surge climate projection efforts (SurgeMIP), Weather Clim. Extrem., 45, https://doi.org/10.1016/j.wace.2024.100689, 2024.

Besio, G., Briganti, R., Romano, A., Mentaschi, L., and De Girolamo, P.: Time clustering of wave storms in the Mediterranean Sea, Nat. Hazards Earth Syst. Sci., 17, 505–514, https://doi.org/10.5194/nhess-17-505-2017, 2017.

Borzì, A. M., Minio, V., De Plaen, R., Lecocq, T., Cannavò, F., Ciraolo, G., D'Amico, S., Re, C. Lo, Monaco, C., Picone, M., Scardino, G., Scicchitano, G., and Cannata, A.: Long-term analysis of microseism during extreme weather events: Medicanes and common storms in the Mediterranean Sea, Sci. Total Environ., 915, https://doi.org/10.1016/j.scitotenv.2024.169989, 2024.

Boumis, G., Moftakhari, H. R., and Moradkhani, H.: A metastatistical frequency analysis of extreme storm surge hazard along the US coastline, Coast. Eng. J., 66, 380–394, https://doi.org/10.1080/21664250.2024.2338323, 2024.

Breilh, J. F., Bertin, X., Chaumillon, É., Giloy, N., and Sauzeau, T.: How frequent is storm-induced flooding in the central part of the Bay of Biscay?, Glob. Planet. Change, 122, 161–175, https://doi.org/10.1016/j.gloplacha.2014.08.013, 2014.

Caruso, M. F. and Marani, M.: Extreme-coastal-water-level estimation and projection: a comparison of statistical methods, Nat. Hazards Earth Syst. Sci., 22, 1109–1128, https://doi.org/10.5194/nhess-22-1109-2022, 2022.

Chowdhury, J. U., Stedinger, J. R., and Lu, L.-H: Goodness-of-fit tests for regional generalized extreme value flood distributions, Water Resour. Res., 27, 1765–1776, https://doi.org/10.1029/91WR00077, 1991.

Cid, A., Menéndez, M., Castanedo, S., Abascal, A. J., Méndez, F. J., and Medina, R.: Long-term changes in the frequency, intensity and duration of extreme storm surge events in southern Europe, Clim. Dyn., 46, 1503–1516, https://doi.org/10.1007/s00382-015-2659-1, 2016.

Coles, S.: An Introduction to Statistical Modeling of Extreme Values, Springer London, London, https://doi.org/10.1007/978-1-4471-3675-0, 2001.

Coles, S. and Casson, E.: Extreme value modelling of hurricane wind speeds, Struct. Saf., 20, 283–296, 1998.

Coles, S. and Tawn, J.: Bayesian modelling of extreme surges on the UK east coast, Philos. T. Roy. Soc. A, 363, 1387–1406, https://doi.org/10.1098/rsta.2005.1574, 2005.

Conte, D. and Lionello, P.: Storm Surge Distribution Along the Mediterranean Coast: Characteristics and Evolution, Procedia Soc. Behav. Sci., 120, 110–115, https://doi.org/10.1016/j.sbspro.2014.02.087, 2014.

Davison, A. C. and Smith, R. L.: Models for Exceedances over High Thresholds, J. R. Statist. Soc. B, 52, 393–442, 1990.

Dias, J. M., Pereira, F., Picado, A., Lopes, C. L., Pinheiro, J. P., Lopes, S. M., and Pinho, P. G.: A comprehensive estuarine hydrodynamics-salinity study: Impact of morphologic changes on Ria de Aveiro (atlantic coast of Portugal), J. Mar. Sci. Eng., 9, 234, https://doi.org/10.3390/jmse9020234, 2021.

Fanjul, A. E., Gomez, B. P., Carretero, J. C., Rodriguez, I., and Arevalo, S.: Tide and surge dynamics along the Iberian Atlantic coast, Oceanol. Acta, 21, 131–143, 1998.

Ferrarin, C., Bellafiore, D., Sannino, G., Bajo, M., and Umgiesser, G.: Tidal dynamics in the inter-connected Mediterranean, Marmara, Black and Azov seas, Prog. Oceanogr., 161, 102–115, https://doi.org/10.1016/j.pocean.2018.02.006, 2018.

Fisher, R. A. and Tippett, L. H. C.: Limiting forms of the frequency distribution of the largest or smallest member of a sample, Math. Proc. Cambridge, 24, 180–190, https://doi.org/10.1017/S0305004100015681, 1928.

Fouad, S. S., Heggy, E., Amrouni, O., Hzami, A., Nijhuis, S., Mohamed, N., Saleh, I. H., Jomaa, S., Elsheshtawy, Y., and Weilacher, U.: Soaring Building Collapses in Southern Mediterranean Coasts: Hydroclimatic Drivers & Adaptive Landscape Mitigations, Earths Future, 13, https://doi.org/10.1029/2024EF004883, 2025.

Gao, H., Shao, Z., Wu, G., and Li, P.: Study of directional declustering for estimating extreme wave heights in the Yellow Sea, J. Mar. Sci. Eng., 8, https://doi.org/10.3390/JMSE8040236, 2020.

Gnedenko, P. B.: Sur la distribution limite du terme maximum d'une serie aleatoire, Ann. Math., 44, 423–453, 1943.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R.: Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form, Water Resour. Res., 15, 1049–1054, https://doi.org/10.1029/WR015i005p01049, 1979.

Haigh, I., Nicholls, R., and Wells, N.: Assessing changes in extreme sea levels: Application to the English Channel, 1900–2006, Cont. Shelf Res., 30, 1042–1055, https://doi.org/10.1016/j.csr.2010.02.002, 2010.

Hamdi, Y., Bardet, L., Duluc, C.-M., and Rebour, V.: Extreme storm surges: a comparative study of frequency analysis approaches, Nat. Hazards Earth Syst. Sci., 14, 2053–2067, https://doi.org/10.5194/nhess-14-2053-2014, 2014.

Hosking, J. R. M.: L-moments: Analysis and Estimation of Distributions using Linear Combinations of Order Statistics, J. Roy. Stat. Soc. B, 52, 105–124, 1990.

Ibáñez, C. and Caiola, N.: Sea-level rise, marine storms and the resilience of Mediterranean coastal wetlands: Lessons learned from the Ebro Delta, Mar. Freshw. Res., 73, 1246–1254, https://doi.org/10.1071/MF21140, 2021.

Jenkinson, A. F.: The frequency distribution of the annual maximum (or minimum) values of meteorological elements, Q. J. Roy. Meteor. Soc., 81, 158–171, https://doi.org/10.1002/qj.49708134804, 1955.

Kirezci, E., Young, I. R., Ranasinghe, R., Muis, S., Nicholls, R. J., Lincke, D., and Hinkel, J.: Projections of global-scale extreme sea levels and resulting episodic coastal flooding over the 21st Century, Sci. Rep., 10, https://doi.org/10.1038/s41598-020-67736-6, 2020.

Kirezci, E., Young, I. R., Ranasinghe, R., Lincke, D., and Hinkel, J.: Global-scale analysis of socioeconomic impacts of coastal flooding over the 21st century, Front. Mar. Sci., 9, https://doi.org/10.3389/fmars.2022.1024111, 2023.

Krit, M., Gaudoin, O., and Remy, E.: Goodness-of-fit tests for the Weibull and extreme value distributions: A review and comparative study, Commun. Stat. Simul. Comput., 50, https://doi.org/10.1080/03610918.2019.1594292, 2021.

Marani, M. and Ignaccolo, M.: A metastatistical approach to rainfall extremes, Adv. Water Resour., 79, 121–126, https://doi.org/10.1016/j.advwatres.2015.03.001, 2015.

Marcos, M., Tsimplis, M. N., and Shaw, A. G. P.: Sea level extremes in southern Europe, J. Geophys. Res.-Oceans, 114, https://doi.org/10.1029/2008JC004912, 2009.

Marra, F., Nikolopoulos, E. I., Anagnostou, E. N., and Morin, E.: Metastatistical Extreme Value analysis of hourly rainfall from short records: Estimation of high quantiles and impact of measurement errors, Adv. Water Resour., 117, 27–39, https://doi.org/10.1016/j.advwatres.2018.05.001, 2018.

Martín, A., Wahl, T., Enriquez, A. R., and Jane, R.: Storm surge time series de-clustering using correlation analysis, Weather Clim. Extrem., 45, https://doi.org/10.1016/j.wace.2024.100701, 2024.

Martins, E. S. and Stedinger, J. R.: Generalized maximum likelihood Pareto-Poisson estimators for partial duration series, Water Resour. Res., 37, 2551–2557, https://doi.org/10.1029/2001WR000367, 2001.

MedECC: Climate and Environmental Change in the Mediterranean Basin – Current Situation and Risks for the Future, First Mediterranean Assessment Report, edited by: Cramer, W., Guiot, J., and Marini, K., Union for the Mediterranean, Plan Bleu, UNEP/MAP, Marseille, France, 632 pp., ISBN: 978-2-9577416-0-1, 2020.

Međugorac, I., Pasarić, M., and Orlić, M.: Historical sea-level measurements at Bakar (east Adriatic), SEANOE [data set], https://doi.org/10.17882/85171, 2026.

Miniussi, A. and Marani, M.: Estimation of Daily Rainfall Extremes Through the Metastatistical Extreme Value Distribution: Uncertainty Minimization and Implications for Trend Detection, Water Resour. Res., 56, https://doi.org/10.1029/2019WR026535, 2020.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I – A discussion of principles, J. Hydrol., 10, 282–290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

Pan, X., Rahman, A., Haddad, K., and Ouarda, T. B. M. J.: Peaks-over-threshold model in flood frequency analysis: a scoping review, Stoch. Env. Res. Risk A., 36, https://doi.org/10.1007/s00477-022-02174-6, 2022.

Pickands, J.: Statistical inference using extreme order statistics, Ann. Stat., 3, 119–131, 1975.

Potisomporn, P., Adcock, T. A. A., and Vogel, C. R.: Extreme value analysis of wind droughts in Great Britain, Renew. Energy, 221, https://doi.org/10.1016/j.renene.2023.119847, 2024.

Ribeiro, A., Barbosa, S. M., Scotto, M. G., and Donner, R. V.: Changes in extreme sea-levels in the Baltic Sea, Tellus A, 66, 20921, https://doi.org/10.3402/tellusa.v66.20921, 2014.

Seneviratne, S. I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vicente-Serrano, S. M., Wehner, M., and Zhou, B.: Weather and Climate Extreme Events in a Changing Climate, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1513–1766, https://doi.org/10.1017/9781009157896.013, 2021.

Serinaldi, F. and Kilsby, C. G.: Rainfall extremes: Toward reconciliation after the battle of distributions, Water Resour. Res., 50, 336–352, https://doi.org/10.1002/2013WR014211, 2014.

Sithara, S., Pramada, S. K., and Thampi, S. G.: Quantifying uncertainty in future sea level projections downscaled from CMIP5 global climate models, Stoch. Env. Res. Risk A., 38, 2065–2079, https://doi.org/10.1007/s00477-024-02669-4, 2024.

Smirlis, Y., Tsioumas, V., Papadimitriou, S., and Tzannatos, E.: An extreme value analysis framework for the tanker market, WMU Journal of Maritime Affairs, https://doi.org/10.1007/s13437-025-00380-x, 2025.

Smith, R.: Extreme value statistics in meteorology and the environment, in: Environmental statistics, Department of Statistics, University of North Carolina, Chapel Hill, NC, USA, 300–377, 2001.

Smith, R. L.: Extreme Value Analysis of Environmental Time Series: An Application to Trend Detection in Ground-Level Ozone, Stat. Sci., 4, 367–377, 1989.

Solari, S., Egüen, M., Polo, M. J., and Losada, M. A.: Peaks Over Threshold (POT): A methodology for automatic threshold estimation using goodness of fit p-value, Water Resour. Res., 53, 2833–2849, https://doi.org/10.1002/2016WR019426, 2017.

Sousa, P. M., Trigo, R. M., Aizpurua, P., Nieto, R., Gimeno, L., and Garcia-Herrera, R.: Trends and extremes of drought indices throughout the 20th century in the Mediterranean, Nat. Hazards Earth Syst. Sci., 11, 33–51, https://doi.org/10.5194/nhess-11-33-2011, 2011.

Tebaldi, C., Ranasinghe, R., Vousdoukas, M., Rasmussen, D. J., Vega-Westhoff, B., Kirezci, E., Kopp, R. E., Sriver, R., and Mentaschi, L.: Extreme sea levels at different global warming levels, Nat. Clim. Chang., 11, 746–751, https://doi.org/10.1038/s41558-021-01127-1, 2021.

Toomey, T., Amores, A., Marcos, M., and Orfila, A.: Coastal sea levels and wind-waves in the Mediterranean Sea since 1950 from a high-resolution ocean reanalysis, Front. Mar. Sci., 9, https://doi.org/10.3389/fmars.2022.991504, 2022.

Toomey, T., Lira-Loarca, A., Marcos, M., Besio, G., and Orfila, A.: Future Wave Climate in the Mediterranean Sea and Associated Uncertainty From an Ensemble of 31 GCM-RCM Wave Simulations, Earths Future, 13, https://doi.org/10.1029/2024EF004992, 2025.

Tsimplis, M. N. and Shaw, A. G. P.: Seasonal sea level extremes in the Mediterranean Sea and at the Atlantic European coasts, Nat. Hazards Earth Syst. Sci., 10, 1457–1475, https://doi.org/10.5194/nhess-10-1457-2010, 2010.

UNEP/MAP: State of the environment and development in the mediterranean, United Nations Environment Programme/Mediterranean Action Plan (UNEP/MAP), ISBN 978-92-807-3796-7, 2020.

Valle-Levinson, A., Marani, M., Carniello, L., D'Alpaos, A., and Lanzoni, S.: Astronomic link to anomalously high mean sea level in the northern Adriatic Sea, Estuar. Coast. Shelf Sci., 257, https://doi.org/10.1016/j.ecss.2021.107418, 2021.

Von Mises, R.: La distribution de la plus grande de n valeurs (reprinted in: Selected Papers of Richard von Mises, Vol. II, American Mathematical Society, Providence, Rhode Island, 271–294, 1964), Rev. Math. Union Interbalcanique, 1, 141–160, 1936.

Vousdoukas, M. I., Voukouvalas, E., Annunziato, A., Giardino, A., and Feyen, L.: Projections of extreme storm surge levels along Europe, Clim. Dyn., 47, 3171–3190, https://doi.org/10.1007/s00382-016-3019-5, 2016.

Woodworth, P. L., Menéndez, M., and Gehrels, W. R.: Evidence for Century-Timescale Acceleration in Mean Sea Levels and for Recent Changes in Extreme Sea Levels, Surv. Geophys., 32, 603–618, https://doi.org/10.1007/s10712-011-9112-8, 2011.

Zorzetto, E., Botter, G., and Marani, M.: On the emergence of rainfall extremes from ordinary events, Geophys. Res. Lett., 43, 8076–8082, https://doi.org/10.1002/2016GL069445, 2016.