Extreme ~~Storm Surge~~ coastal water level estimation and projection ~~through the Metastatistical Extreme Value Distribution~~: A comparison of statistical methods

Maria Francesca Caruso[1] and Marco Marani[1]

[1]Department of Civil, Architectural, and Environmental Engineering, University of Padova, 35131, Padova, Italy

Correspondence: Caruso, M.F. (mariafrancesca.caruso@phd.unipd.it)

Abstract. Accurate estimates of the probability of extreme sea levels are pivotal for assessing risk and the design of coastal defense structures. This probability is typically estimated by modelling observed sea-level records using one of a few statistical approaches. In this study we comparatively apply the Generalized Extreme Value (GEV) distribution, based on Block Maxima (BM) and ~~Peak-Over-Threshold~~ Peaks-Over-Threshold (POT) formulations, and the recently Metastatistical Extreme Value Distribution (MEVD) to four long time series of sea-level observations distributed along European coastlines. A cross-validation approach, dividing available data in separate calibration and test sub-samples, is used to compare their performances in high-quantile estimation. To address the limitations posed by the length of the observational time series, we quantify the estimation uncertainty associated with different calibration sample sizes, from 5 to 30 years. Focusing on events with a high return period, we find that the GEV-based approaches and MEVD perform similarly when considering short samples (5 years), while the MEVD estimates outperform the traditional methods when longer calibration sample sizes (10-30 years) are considered. We then investigate the influence of sea-level rise through 2100 on ~~storm surges~~ extreme sea-level frequencies. The projections indicate an increase in the height of ~~storm surges~~ total water levels for a fixed return period that are spatially heterogeneous across the coastal locations explored.

1 Introduction

The statistical analysis of extreme values of random variables is of wide conceptual and applicative importance in science and engineering (Coles, 2001; Beirlant et al., 2004; Castillo et al., 2005; Finkenstädt and Rootzén, 2004). Modelling extreme value probability of occurrence is indeed an active field of theoretical and applied research in many fields, such as hydrology and climatology ~~(Katz et al., 2002)~~ ~~, ecology (Katz et al., 2005)~~ (Katz et al., 2002; Cancelliere, 2017; Mekonnen et al., 2021; Miniussi and Marra, 2021), ecology (Katz et al., 2005; Rypkema et al., 2019), ocean wave modelling (Rueda et al., 2016; Benetazzo et al., 2017; Barbariol et al., 2019), transport engineering (Songchitruksa and Tarko, 2006), geophysical processes ~~(Pisarenko et al., 2014a, b; Hosseini et al., 2020)~~ (Pisarenko et al., 2014a, b; Elvidge and Angling, 2018; Hosseini et al., 2020), biomedical data analysis (De Zea Bermudez and Mendes, 2012; Chiu et al., 2018), insurance and financial applications ~~(Embrechts et al., 1997)~~ (Embrechts et al., 1997; Chan et al., 2022), and many others.

In particular, the reliable estimation of the occurrence probability of coastal flooding events of large magnitude is crucial to environmental hazard evaluation (Coles and Tawn, 2005; Hamdi et al., 2018) and to decision-making and mitigation measure design. In fact, coastal flooding hazard has been increasing at the global scale in recent decades, a trend expected to continue as a result of climate change (Meehl et al., 2007; Church et al., 2013; Fortunato et al., 2016). Several studies highlight that global sea-level rise will continue accelerating in the 21st century as a consequence of climate change (Church and White, 2006; Jevrejeva et al., 2008; Church and White, 2011; Haigh et al., 2014b; Hay et al., 2015). Additionally, changes in storminess may have an important role in modifying the frequency and magnitude of water level extremes (Lowe et al., 2010; Menéndez and Woodworth, 2010; Woodworth et al., 2011). Much of the current work on extreme coastal flooding events is based on the classical Extreme Value Theory (EVT) ~~(Fréchet, 1927; Gumbel, 1935, 1958; Gnedenko, 1943; Chow, 1953; Dalrymple, 1960; Pickands, 1975; Hosking and Wallis, 1987, 1993,~~ (Fréchet, 1927; Dalrymple, 1960; Coles, 2001; Woodworth and Blackman, 2002; Hamdi et al., 2014, 2015, and references therein), which identifies the family of distribution functions known as Generalized Extreme Value (GEV) distribution (Von Mises, 1936) as a general model for the distribution of maxima (or minima) extracted from fixed time periods of equal length ("blocks", most commonly with length of one year). The GEV, according to its original formulation, arises as a limiting distribution for maxima (or minima, not considered here) of a sequence of independent and identically distributed (i.i.d.) random variables. The ~~Peak-Over-Threshold~~ Peaks-Over-Threshold (POT) formulation (Balkema and de Haan, 1974; Pickands, 1975), extends the original GEV derivation by modelling all events exceeding a high threshold, as opposed to considering just yearly maxima as in the GEV-Block Maxima formulation (GEV-BM). The POT approach again recovers the GEV distribution as the distribution of the annual maxima if two assumptions are valid (Davison and Smith, 1990): 1) the number of events/year is Poisson-distributed; 2) exceedances over the threshold come from a Generalized Pareto Distribution (GPD). Under these suitable conditions, in the following we will refer to the POT framework as POT-GPD formulation. For a brief overview of the theory underlying EVT and ~~POT method~~ the two main methods based on the GEV distribution (i.e. BM and POT approaches), the reader can refer to ~~supplementary material. The GEV-POT~~ the Method section or the supplementary material for more details. The POT-GPD approach is often considered to be superior to GEV-BM in practical applications, due to its more efficient use of often scarce observations. For extreme sea-level studies in particular, Coles and Tawn (2005) and Haigh et al. (2010) recognize two weakness in the use of the GEV-BM analysis: 1) sea level is the combination of tide-driven (deterministic) and storm-driven (stochastic) components. The presence of a deterministic component is suggested to violate the i.i.d. assumption required in the GEV-BM derivation; 2) sea-level data are collected frequently (e.g., hourly), while the GEV-BM approach only studies annual maxima, with an extremely inefficient use of the data. The POT framework exploits more of the available information with respect to the BM approach (e.g., Coles, 2001; Bernardara et al., 2014). However, the choice of a suitable threshold to retain a few above-threshold events/year is a critical step, and the estimation uncertainty significantly depends on threshold selection ~~(Önöz and Bayazit, 2001; Li et al., 2012)~~ (Önöz and Bayazit, 2001; Li et al., 2012; Solari et al., 2017). The selected threshold value implies a balance between bias and estimation error variance (Coles, 2001). In fact, too low a threshold will violate the independence hypothesis of the framework, leading to bias, while too high a threshold will retain just a few values above the threshold, leading to high error variance.

More generally, GEV-based approaches, by construction, discard most of the observations, and do not attempt to optimize the use of the available information (Volpi et al., 2019). Furthermore, the traditional extreme value theory derives the GEV distribution either as the asymptotic distribution when the number of events/block becomes very large, or through the ad-hoc GPD-Poisson assumptions underlying the POT approach. Whether these hypotheses do apply to the case of storm surges is a matter of discussion, but it seems beneficial to adopt methods that require the least amount of a-priori assumptions on the properties of the event arrival process. As a contribution to overcoming the limitations of the traditional EVT, here we explore the use of an alternative approach for modelling extreme sea levels, the Metastatistical Extreme Value Distribution (MEVD). This approach was introduced by Marani and Ignaccolo (2015) and has been previously applied to rainfall, flood-frequency analysis, and hurricane intensities. The MEVD models the distribution of yearly maxima starting from the distribution of "ordinary values", i.e. all the available data, in contrast to just considering annual maxima or a few values above a threshold. Moreover, the MEVD framework (i) is a non-asymptotic extreme value distribution, which does not require the number of events/year to be large as in the traditional theory, and (ii) makes no a-priori assumptions on the properties of the event occurrence process (while, e.g. POT-GPD assumes a Poisson occurrence process). In previous applications, the MEVD has been shown to significantly reduce estimation uncertainty compared to traditional approaches, especially when considering return periods greater than the sample size used for parameter estimation (Zorzetto et al., 2016; Marra et al., 2018; Miniussi and Marani, 2020; Miniussi et al., 2020a, b).

Here we comparatively analyze the performance of GEV-based approaches and ~~of the~~ MEVD in high-quantile estimations with application to extreme sea levels at different observation sites. The aim is to: 1) identify the statistical tool affording minimal uncertainty in the estimate of extreme sea levels with assigned probability of exceedance, and 2) model and understand how climate change will affect the extreme sea-level occurrence. To achieve these objectives, we analyze selected sea level time series along the European coastline and evaluate extreme sea level predictive uncertainty by adopting a cross-validation approach, in which calibration and test samples are kept separate and independent. Subsequently, we use the optimized estimation method to infer possible changes in coastal flooding hazard under IPCC climate change scenario RCP4.5 and RCP8.5.

The structure of the paper is as follows: Section 2 outlines the sea-level data and the methodology used in this application, results are described in Section 3, while the conclusions are given in Section 4.

2 Materials and Methods

2.1 Data

The analyses were performed using daily and hourly sea-level records from four tide gauge stations (see Table 1) distributed along European coastlines: Venice (Italy), Hornbæk (Denmark), Marseille (France), and Newlyn (United Kingdom). The study sites span a variety of geographical locations, coastal morphologies and storm regimes.

Venice sea-level data (maximum and minimum daily observations) were obtained from the "Centro Previsioni e Segnalazioni Maree" of the Venice Municipality (https://www.comune.venezia.it/it/content/centro-previsioni-e-segnalazioni-maree) for the

Table 1. Information of sea-level data used in this application.

Site name	Country	Location (degree, min.) Lat.	Long.	Period	Missing years (%)	Deleted years	Number of years
Venice	Italy	45°25.0'N	12°20.0'E	1872-2019	-	-	148
Hornbæk	Denmark	56°06.0'N	12°28.0'E	1891-2012	-	1985	121
Marseille	France	43°16.7'N	5°21.2'E	1885-2018	14.2	1897; 1918; 1919; 1928; 1937; 1940; 1998; 2009; 2010	106
Newlyn	United Kingdom	50°06.1'N	5°32.5'W	1915-2016	-	1984; 2010	100

Punta della Salute gauge station . The remaining water level data, all at the hourly scale, were downloaded from the University of Hawaii Sea Level Center (UHSLC) repository (http://uhslc.soest.hawaii.edu/data/?rq#uh745a/).

95 All sea-level datasets span long observational periods: 148 years for Venice, 122 years for Hornbæk, 115 years for Marseille (ca. 19 missing years) and 102 years for Newlyn.

The raw data for all stations were pre-processed to eliminate: 1) years with less than six months of water level observations, and 2) days with less than 24 h of data (for the case of hourly data). This process yields four "cleaned up" time series that were subsequently used in the analyses (see Table 1). Figure 1 shows daily maximum sea levels at the gauge stations explored after

100 pre-processing.

2.2 Methods

2.2.1 Mean sea level removal

The sea-level sequence is highly correlated and is generated by a non-stationary process due to long term trends in mean sea level, the deterministic tidal component, surge seasonality, and interactions between the tide and surge

105 (Dixon and Tawn, 1999). Tide-surge interactions may change amplitude and phase of the surges, mostly in shallow estuarine areas (Johns and Ali, 1980; Bernier and Thompson, 2007; Zhang et al., 2010). However, this effect needs to be taken into account when separating the surge and tide components. Here, we do not attempt to separate these contributions, but we only analyze the sum given by the combination of the water level setup, induced by meteorological forcing, and the astronomical tide. Hence, we simply study such sum as the final result of the non-linear interactions between individual components. Under

110 this premise, for a given site and at any instant of time t, the observed sea level $z(t)$ (after averaging out waves), can be split into three components (Pugh and Vassie, 1979): mean sea level, $msl(t)$, astronomically induced tidal level, $x(t)$, and meteorologically induced surge level, $y(t)$:

$$z(t) = msl(t) + x(t) + y(t) \tag{1}$$
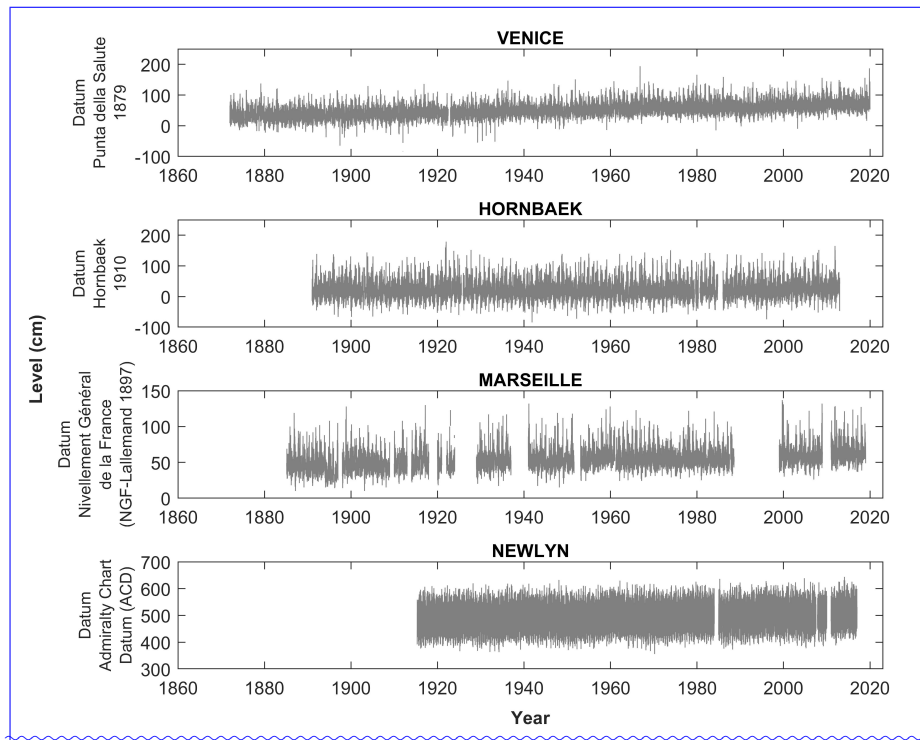
Figure 1. Daily maximum sea levels at different gauge stations explored after pre-processing: Venice (IT), Hornbæk (DK), Marseille (FR), and Newlyn (UK).

$msl(t)$ represents the long-term variations of water levels and of the elevation datum (i.e. possible land subsidence or uplift). Local $msl(t)$ does not change uniformly over time and its calculation is affected by many factors, such as tidal phases,

115 long-term wind and atmospheric pressure patterns, vertical land motion (subsidence or uplift). The tidal contribution to the instantaneous sea level, $x(t)$, caused by the gravitational forces exerted by the moon and the sun is deterministic in nature, and can be predicted with a good degree of accuracy. This tidal variability occurs with characteristic periodicities between 12 hours and 18.61 years (Eliot, 2010; Haigh et al., 2011; Pugh and Woodworth, 2014; Peng et al., 2019; Valle-Levinson et al., 2021). This latter longest tidal periodicity corresponds to the precession of the lunar nodal cycle. The storm-surge contribution,

120 $y(t)$, is the meteorologically-induced change in the water level generated by a combination of factors, such as the magnitude and direction of the wind, spatial gradients in atmospheric pressure, storm size, fetch, bathymetry, storm duration, etc (Hall and Sobel, 2013).

Two classes of methods are widely used to estimate the probability of occurrence of extreme sea levels: direct and indirect methods. Indirect methods model separately the deterministic and the stochastic components of $z(t)$ then recombined by

125 convolution. Examples are the joint probability method (Pugh and Vassie, 1979, 1980), the revised joint probability method (Tawn and Vassie, 1989), the exceedance probability method (Middleton and Thompson, 1986; Hamon and Middleton, 1989) , and the empirical simulation technique (Scheffner et al., 1996; Goring et al., 2011). Direct methods, such as the one adopted

here, analyze observed values compounding the astronomical and stochastic storm-surge component. Direct methods mostly differ based on the analysis approach adopted, such as the annual maxima method (Jenkinson, 1955; Gumbel, 1958), the peaks-over-threshold method (Davison and Smith, 1990), or the r-largest method (Smith, 1986; Tawn, 1988). Here, we study the distribution of the sum, $h(t)$, of the contributions from the deterministic tide and the stochastic surge:

$$h(t) = z(t) - msl(t) \tag{2}$$

From a statistical point of view, this choice is justified by the fact that the random arrival of storms adds a stochastic surge contribution at unpredictable times, thereby causing $h(t)$ to be values from a random variable, even though it contains a deterministic component. The presence of a deterministic component of course does imply a strong auto-correlation in the observed signal, which will be subsequently filtered out by suitable signal processing described below.

$msl(t)$, is here computed as the yearly average of daily levels. The yearly average is chosen, rather than the customary 19-year average that eliminates all tidal periodicities, however small in amplitude, to better capture the surge contribution that causes the water level to deviate during a storm with respect to the "current", yearly, value of $msl(t)$. Once $h(t)$ is computed by removing $msl(t)$ from recorded levels, all local maxima of $h(t)$, or water level peaks, are identified and their values constitute the basis for subsequent analyses.

In the following discussion, we will use the terms "total water level" and "coastal water level" when referring to the quantities $z(t)$ and $h(t)$ respectively.

2.2.2 Extreme Value Theory

As highlighted by Serinaldi and Kilsby (2014), the EVT deals with the asymptotic distributional behaviour of two types of data modeled with two well-known approaches, namely the so-called Block Maxima (BM) and Peaks-Over-Threshold (POT). The first type models the maximum values extracted from blocks of fixed length, whereas the second one models all the exceedances of high threshold. The cornerstones of the EVT are two theorems: the Fisher-Tippett-Gnedenko theorem (also known as the three types theorem, Fisher and Tippett (1928); Gnedenko (1943); Gumbel (1958)) and the Pickands-Balkema-de Haan theorem (also known as the second theorem of EVT, Balkema and de Haan (1974); Pickands (1975)).

According to the three type theorem, there are three possible non-degenerate distribution functions which can arise as limiting distributions of extremes of random sample: (i) the Gumbel distribution or type I, (ii) the Fréchet distribution or type II, and (iii) the reverse-Weibull distribution or type III. The above three limiting distribution laws can be combined into a single family of three-parameter distribution known as the Generalized Extreme Value (GEV) distribution given by:

$$G(x; \mu, \psi, \xi) = exp\{-[1 + \frac{\xi}{\psi} \cdot (x - \mu)]\}^{-1/\xi} \tag{3}$$

defined on the region for which $\{x : 1 + \frac{\xi}{\psi} \cdot (x - \mu) > 0\}$. In Eq. 3 $\mu \in (-\infty, +\infty)$ is a location parameter, $\psi > 0$ is a scale parameter, and $\xi \in (-\infty, +\infty)$ is a shape parameter which controls the nature of the tail distribution (Frechét type for $\xi > 0$, Gumbel type for $\xi = 0$, and reverse-Weibull type for $\xi < 0$).

The second theorem of EVT defines a method to model the tail of distribution above a threshold value (Davison and Smith, 1990). In particular, the theorem states that for a large enough threshold value, u, the distribution of exceedances of some high threshold ($y = X - u$, where X is a sequence of i.i.d. random variables), is described by a Generalized Pareto Distribution (GPD), which has the following cumulative distribution function:

$$G(y; \sigma_u, \xi) = 1 - (1 + \frac{\xi}{\sigma_u} \cdot y)^{-1/\xi} \tag{4}$$

defined on $\{y : y > 0 \text{ and } (1 + \frac{\xi}{\sigma_u} \cdot y > 0)\}$, where σ_u and ξ are the shape and scale parameters respectively.

There is a link between these two distributions according to which if block maxima have approximate distribution GEV, then threshold excesses have corresponding approximate distribution within the generalized Pareto family and vice versa, GEV can be obtained from GPD under two appropriate conditions (i.e., the occurrences are Poisson-distributed and excesses over threshold come from a GPD). The duality between Eq. 3 and Eq. 4 means that the GPD parameters of the excesses are uniquely determined by those of the associated GEV distribution of block maxima (see e.g, Coles (2001)). In particular, the shape parameter, ξ, is equal to that of the corresponding GEV distribution and the scale parameters of the two distributions are related by $\sigma_u = \sigma + \xi \cdot (u - y)$.

The interested reader can refers to Coles (2001) for a detailed description of statistical methods for extremes in hydrology or Papalexiou and Koutsoyiannis (2013) for a recent overview of the history of the EVT.

2.2.3 The Metastatistical Extreme Value Distribution

The typical EVT derivation starts from the premise that the maximum value among n realizations of a random variable (M_n) is distributed according to the cumulative distribution function $P(X \leq x) = G(x) = F(x; \overrightarrow{\theta})^n$ $P(M_n \leq x) = G(x) = F(x; \overrightarrow{\theta})^n$ (where, as customary, a capital letter indicates the random variable and a small cap letter indicates a value of the random variable). This approach assumes that the n values of the random variable of interest are generated by the same distribution, the "ordinary value" distribution $F(x; \overrightarrow{\theta})$, and are thus independent and identically distributed. n is the number of events in a block, such that $G(x)$ is the cumulative distribution of the block maxima. The classical EVT assumes that either the number of events/year block is large (asymptotic hypothesis, leading to the GEV-BM formulation) or that the number of events/year block above a high threshold is distributed according to a Poisson distribution (GEV-POT POT-GPD formulation). The recently-proposed Metastatistical Extreme Value Distribution (Marani and Ignaccolo, 2015) is a doubly stochastic approach (Dubey, 1968; Beck and Cohen, 2003) that relaxes these hypotheses by considering both the parameters ($\overrightarrow{\theta}$) of the ordinary value probability distribution and the number of events/year block to be random variables. The Hence, the MEVD cumulative distribution of block maxima (estimated using a much greater sample than just yearly maxima used in the BM approach) is then defined as the compound probability:

$$G(x) = \sum_{n=1}^{+\infty} \int_{\Omega_{\overrightarrow{\Theta}}} F(x; \overrightarrow{\theta})^n g(n, \overrightarrow{\theta}) d\overrightarrow{\theta} \tag{5}$$

where $g(n, \vec{\theta})$ is the joint probability distribution of the number of events in a ~~year~~ block and of the parameters vector (discrete in N and continuous in $\vec{\Theta}$), $\Omega_{\vec{\Theta}}$ is the population of all possible parameter values.

For practical applications, the MEVD can be approximated by substituting ensemble average in Eq. 5 with the sample average computed over all the blocks in the time series, obtaining:

$$G(x) \cong \frac{1}{M} \sum_{j=1}^{M} F(x; \theta_j)^{n_j} \tag{6}$$

where M is the number of blocks in the historical record, $F(x; \vec{\theta_j})$ is the cumulative distribution of ordinary values in the jth ~~year~~block, and n_j is the number of events in the jth ~~year~~block. A common choice for the block length is 1 year. Note that the values of the parameters $\vec{\theta_j}$ may be estimated on Estimation Windows (EW) with length that is different from block length. For example, if the block length is 1 year, it may be advantageous to estimate parameter values on longer time slices to ensure, depending on the rate of event occurrence, that a reliable estimation of the parameters is possible. Miniussi and Marani (2020) in applications to daily rainfall extremes find that, when the number of events per year is less than 20-25 events/years, then the optimal EW length may be greater than one year.

It is interesting to note that the POT approach, briefly described above, can be thought of as a particular case of MEVD. In fact, Zorzetto et al. (2016) highlight that if one assumes (i) x to be the excess over a high threshold, (ii) $F(x; \vec{\theta_j})$ to be a Generalized Pareto Distribution (with fixed, deterministic parameters), and (iii) n to be generated by a Poisson distribution, then the GEV distribution is recovered as a particular case of the MEVD by means of the POT approach.

MEVD has been applied in several earth-science contexts. In rainfall extremes estimates, the ordinary value distribution is assumed to be Weibull when applied to point daily rainfall (Marani and Ignaccolo, 2015; Zorzetto et al., 2016; Schellander et al., 2019; Miniussi and Marani, 2020; Miniussi et al., 2020b), point sub-daily rainfall (Marra et al., 2018), and satellite rainfall estimates (Zorzetto and Marani, 2019, 2020). For flood across the United States, Miniussi et al. (2020a) propose to adopt a Gamma distribution for $F(x; \vec{\theta_j})$. Hosseini et al. (2020) describe Atlantic hurricane intensities using a Generalized Pareto ordinary value distribution. In all cases the appropriate form for the underlying ordinary value distribution was identified by minimizing the estimation uncertainty within a cross-validation approach, which is also followed here. In this particular application to extreme coastal water levels, three candidate probability distributions for $F(x; \vec{\theta_j})$ in Eq. 6 are tested, i.e. the Gamma, Weibull and Generalized Pareto distributions. Based on the comparative evaluation of the performance of these distributions, the Generalized Pareto distribution emerged as the best model for the "ordinary" coastal water level values.

In the present context, we define as ordinary values any storm-surge water elevation (i.e. the maximum water level reached during a storm event) greater than a site-specific threshold value. This threshold is chosen to be as small as possible (differently from the POT approach), to retain as much of the observational information as possible, and will be dependent on the ~~magnitudes of the tidal and~~ magnitude of the local tidal range (sea level difference between high and low water level over a tidal cycle) and of storm contributions. Additionally, the threshold is set to be large enough to filter out coastal water level peaks that are likely fully determined by tidal fluctuation, in the absence of any storm contribution. Given the above constraints, we also choose the threshold value that minimizes the estimation error under the MEVD framework.

As suggested by several rainfall applications, ordinary distribution parameters are here estimated using the Probability
220 Weighted Moments (PWMs) method in non-overlapping estimation windows of 5 years. In the present application, the optimal
estimation window length was set to 5 years to obtain a more robust parameters estimation, especially when few values in each
year are available. PWM estimation, introduced by Greenwood et al. (1979), is widely applied because of its good performance,
particularly in the presence of small sample sizes, its reduced estimation bias and sensitivity to the presence of outliers in the
data (Hosking et al., 1985; Hosking and Wallis, 1987; Hosking, 1990).

225 **2.2.4** ~~**Mean sea level removal**~~

~~The sea-level sequence is highly correlated and is generated by a non-stationary process due to long term trends
in mean sea level, the deterministic tidal component, surge seasonality, and interactions between the tide and surge
(Dixon and Tawn, 1999). Tide-surge interactions may change amplitude and phase of the surges, mostly in shallow
estuarine areas (Johns and Ali, 1980; Bernier and Thompson, 2007; Zhang et al., 2010). However, in deeper sea conditions the~~
230 ~~non-linear interactions between tide and surge have been found to be negligible (Lionello et al., 2006). Hence, we will neglect
this effect here and will consider the observed sea level as the sum of additive components. Under this premise, for a given
site and at any instant of time t, the observed sea level $z(t)$ (after averaging out waves), can be split into three components
(Pugh and Vassie, 1979): mean sea level, $msl(t)$, astronomically induced tidal level, $x(t)$, and meteorologically induced surge
level, $y(t)$:~~

235 $$z(t) = msl(t) + x(t) + y(t)$$

~~$msl(t)$ represents the long-term variations of water levels and of the elevation datum (i.e. possible land subsidence or uplift).
Local $msl(t)$ does not change uniformly over time and its calculation is affected by many factors, such as tidal phases,
long-term wind and atmospheric pressure patterns, vertical land motion (subsidence or uplift). The tidal contribution to the
instantaneous sea level, $x(t)$, caused by the gravitational forces exerted by the moon and the sun is deterministic in nature, and~~
240 ~~can be predicted with a good degree of accuracy. This tidal variability occurs with characteristic periodicities between 12 hours
and 18.61 years (Eliot, 2010; Haigh et al., 2011; Pugh and Woodworth, 2014; Peng et al., 2019; Valle-Levinson et al., 2021).
This latter longest tidal periodicity corresponds to the precession of the lunar nodal cycle. The storm-surge contribution,
$y(t)$, is the meteorologically-induced change in the water level generated by a combination of factors, such as the magnitude
and direction of the wind, spatial gradients in atmospheric pressure, storm size, fetch, bathymetry, storm duration, etc~~
245 ~~(Hall and Sobel, 2013).Two classes of methods are widely used to estimate the probability of occurrence of extreme sea levels:
direct and indirect methods. Indirect methods model separately the deterministic and the stochastic components of $z(t)$. A
direct method is adopted here, in which we study the distribution of the sum, $h(t)$, of the contributions from the deterministic
tide and the stochastic surge:~~

$$h(t) = z(t) - msl(t)$$

250 ~~From a statistical point of view, this choice is justified by the fact that the random arrival of storms adds a stochastic surge
contribution at unpredictable times, thereby causing $h(t)$ to be values from a random variable, even though it contains a~~

~~deterministic component. The presence of a deterministic component of course does imply a strong auto-correlation in the observed signal, which will be subsequently filtered out by suitable signal processing described below.~~ $msl(t)$, ~~is here computed as the yearly average of daily levels. The yearly average is chosen, rather than the customary 19-year average that eliminates~~

255   ~~all tidal periodicities, however small in amplitude, to better capture the surge contribution that causes the water level to deviate during a storm with respect to the "current", yearly, value of $msl(t)$. Once $h(t)$ is computed by removing $msl(t)$ from recorded levels, all local maxima of $h(t)$, or water level peaks, are identified and their values constitute the basis for subsequent analyses.~~

2.2.4   Selection of independent events

260   The GEV-based approaches are fit on either annual peak maxima (GEV-BM) or on a few water level peaks over a high threshold (~~GEV-POT~~POT-GPD), which can be assumed to be realizations of independent stochastic variables. The MEVD requires that all ordinary values (coastal water level peaks in this case) within one block may be assumed to be realizations from independent random variables. This hypothesis, in turn, requires that observed peaks are filtered to only retain events that may be considered to be independent, through a de-clustering process (Coles, 2001; Ferro and Segers, 2003; Beirlant et al., 2004; Bommier, 2014;

265   Marra et al., 2018). Several criteria have been developed for such processing of the data. A common criterion sets the minimal time separation, or lag (τ), for two events to be considered independent. Intuitively, high water-level events separated by a sufficiently long time period are reasonably caused by distinct storm events. However, when analyzing the water level with respect to current mean sea level, a quantity that contains the deterministic tidal contribution, dependence may be expected to be present also for large lags. In theory, some dependence is present for lags up to the longest periodicity in the tidal

270   signal (18.61 years). In practice, as the dependence in the tidal signal decreases for increasing lag, one expects that a much shorter threshold time lag will be sufficient to make sure that only independent events are considered~~(.~~ The analysis of the correlograms of selected coastal water level peaks shows that some correlation persists also for long time lags and also in the de-clustered time series. Even though the strength of this correlation is relatively small (the autocorrelation function, ACF, is always less than 0.3), it could impact the ability of the MEVD, which assumes independence, to capture observed extreme

275   behaviour. The de-clustering process does significantly decrease correlation, as may be seen by comparing Figure S1 ~~).~~(ACF prior to de-clustering) and Figure S2 (after de-clustering). Interestingly, it is seen that the tidal contribution (that generates periodicities in the ACF) is strongly visible in Venice and Newlyn, while it is quite small in Hornbæk and Marseille. The underlying tidally-induced correlation becomes more clearly visible after de-clustering also in Hornbæk and Marseille. We note that the existing literature implementing de-clustering approaches to coastal level signals normally focuses on studying

280   the storm-surge component only. As result, it uses threshold time lag values that are smaller than those adopted here because characteristic correlation times of the surge component are significantly smaller than those associated with the sum given by the combination of surge and tidal components. For example, the independence between two consecutive storm surge events in southern Europe has been found to be achieved with a threshold lag of 3 days (Cid et al., 2015). A threshold separation of one day between consecutive events is imposed by Tebaldi et al. (2012) in their analysis of storm surges along the US coast. Haigh

285   et al. (2010) ~~adopted~~ adopt a threshold lag of 30 h in the English Channel, while Bernardara et al. (2011) ~~adopted~~ assume a 72

h independence criterion. After exploring values between 24 h and several days, we ~~adopted~~ adopt a threshold lag of 30 days, which yielded the minimum estimation error under the MEVD approach, and is consistent with the main lunar periodicity. The result of this ~~declustering~~ de-clustering process is a set of independent events with magnitudes h_k, whose number n_j in year - or block - j is a realization of a random variable as illustrated in Eq.s 5 and 6.

2.2.5 Cross-validation procedure

Statistical modelling aims to use sample information to infer the probability distribution of the population from which the data are extracted. This inference is uncertain due to imperfect parameter estimates and to the possible inability of the chosen distribution to capture the statistical properties of the underlying population. Although these sources of uncertainty are inherent in any statistical model, their impact can be minimized by a careful choice of the model and by an effective use of all sources of information (Coles, 2001). In many applications uncertainty is estimated by means of goodness-of-fit measures, which quantify how well the distribution compares to the sample on which it was fitted. However, this procedure does not provide a measure of the predictive uncertainty encountered when trying to estimate the probability of occurrence of the "next", yet unobserved value. In this study, we evaluate the performance in high-quantile estimation associated with the use of the MEVD and the GEV distribution, by means of a cross-validation (CV) procedure, in which model predictions of the probability of occurrence are compared to frequencies from data that were not used in the estimation of model parameters. This is possible by dividing observations into two sets of independent data: the estimation set is the sample from which model parameters are estimated and the test set is the sample with which model predictions are compared.

The procedure can be summarized as follows: a) we randomly reshuffle the observational years on record while keeping all the water level independent peaks in their original year to 1) preserve both the ordinary value frequency distribution in each year and the distribution of the number of events/year, and 2) remove possible non-stationarity and correlation in the time series; b) we divide the observational sample into two independent sub-samples obtained by randomly selecting S years from the original time series of length M: this sub-sample (in the following "calibration sample") is used for parameter estimation, while data in the remaining $V = M - S$ years are used for testing (~~validation sample~~ in the following "validation - or test - sample"); c) ~~we~~ as usual in frequency analysis, we associate to each observed yearly maximum, x_i, an empirical frequency value given by Weibull's estimator $F_i = i/(V + 1)$ where i is the rank of x_i in the list of yearly maxima sorted in ascending order, and $V = M - S$ is the sample size in the validation sub-sample. The return period Tr associated with each yearly maximum is then simply $Tr_i = 1/(1 - F_i)$; d) we estimate the GEV and MEVD quantiles using the parameter values estimated in step b) from the calibration sub-sample; ~~d)~~ e) focusing on the validation sub-sample, in every realization (for $p = 1, \ldots, Nr$; $Nr = 1000$ here) and for a fixed mean recurrence time (Tr), we compute the Non Dimensional Error between the estimated and observed quantiles as follows: $NDE_p(S, Tr) = [h_{(est,p)}(S, Tr) - h_{(obs,p)}(S, Tr)]/h_{(obs,p)}(S, Tr)$; ~~e~~f) we repeat the CV scheme above Nr times. This procedure is performed for different calibration sample sizes (S = 5, 10, 20, and 30 years) to evaluate how estimation uncertainty varies with return period and calibration sample size.

2.2.6 Future ~~storm-surge~~ total water level projections

Future increases in the frequency of ~~intense storm surges~~ extreme total water levels (i.e. the variable previously referred as $z(t)$) due to climate change will have serious impacts on coastal regions. These impacts will vary temporally and regionally, depending on ~~the local~~ (i) the local relative mean sea-level rise (including possible subsidence or uplift), (ii) current storm-surge intensity probability distributions, and (iii) changes in the dominant meteorological dynamics. In this particular application to extreme ~~storm surges,~~ coastal water levels (i.e. the sum given by the combination of the water level setup, induced by meteorological forcing, and the astronomical tide), only the first two factors are considered.

It is very likely that sea-level rise will continue to accelerate over time, thereby increasing the frequency of extreme sea level events, leading to severe flooding in many low-lying coastal cities and small islands (Oppenheimer et al., 2019). ~~As Haigh et al. (2011) underline, changes in mean sea level have an effect on extreme sea levels in two ways: (i) directly: a rise (or fall) in mean sea level will result in a lower (or higher) surge elevation at high tide being necessary to produce a sea level high enough to cause flooding; (ii) indirectly: changes in mean sea level alter water depths~~ Various techniques have been used to study possible changes in coastal flooding hazard (e.g., McInnes et al., 2013; Vousdoukas et al., 2016). Several authors ~~(Zhang et al., 2000; Woodworth and Blackman, 2004)~~ have found that past variations in the frequency of occurrence of extreme sea levels have been primarily determined by changes in mean sea level ~~. This suggests that indirect effects and~~ (e.g., Zhang et al., 2000; Woodworth and Blackman, 2004; Lowe et al., 2010; Menéndez and Woodworth, 2010; Haigh et al., 2014b; Wahl et al., 2017) . This implies that effects of variations in storminess (e.g., magnitude, trajectories and frequency) have been small ~~,~~ in the observational record, compared ~~with the direct effects (Haigh et al., 2014a). We thus focus on the latter here.~~ to the dominant effects of mean sea-level changes (Haigh et al., 2014a). This notion is also confirmed by our trend analyses of maximum yearly departures from the average sea level (see §3.1)), which fail to detect trends in the maximum difference between total sea level and concurrent mean sea level except at one of the sites (Venice), where it is smaller (0.7 mm/yr) than past and projected rates of sea-level rise (respectively $\sim$3.0 mm/yr and $\sim$8.0 mm/yr at the end of the century, according to the RCP8.5 IPCC scenario).

~~Various techniques have been used to study possible changes in coastal flooding hazard (e.g., McInnes et al., 2013; Vousdoukas et al., 2016).In this application,~~ Based on these elements, here we estimate the probability of future ~~storm surges~~ total water levels along European coastlines ~~are estimated~~ by assuming that changes in the tidal and storm-surge components ~~, computed~~ are negligible with respect to ~~the current mean sea level, do not change as mean~~ changes in mean sea-level~~changes. Hence, we do not include the effects of possible changes in local "storminess", a widely utilized approach~~, an assumption common to previous approaches (Araújo and Pugh, 2008; Haigh et al., 2010; Tebaldi et al., 2012).

To assess how the exceedance probabilities of extreme total water levels might change in the future, the projections of sea-level rise through 2100 from the IPCC's Fifth Assessment Report (AR5) are used. In particular, we explore an intermediate (RCP4.5) and an extreme scenario (RCP8.5), using CMIP5 model outputs from the "Integrated Climate Data Center" (ICDC) database (University of Hamburg: https://icdc.cen.uni-hamburg.de/en/ar5-slr.html).

For each tide gauge, our approach can be summarized as follows: 1) we infer the probability distribution of extreme coastal water levels (annual maxima) from observed independent ~~surge~~ events whose intensity (maximum coastal water level attained, h_k) is defined with respect to the concurrent mean sea level computed on a yearly basis; 2) we estimate the future probability of extreme ~~sea~~ total water levels by translating extreme level quantile estimates upward according to the projected mean sea level at each location (thereby implicitly assuming subsidence/uplift to be negligible).

2.2.7   Return period

One of the main objectives of frequency analysis is to calculate the average recurrence interval or return period. It is a widely used concept in hydrological and geophysical risk analysis. If a process is stationary, the return period (Tr) of an event magnitude is defined as the average time elapsing between two consecutive exceedances of this magnitude. Alternatively, it may be said that a magnitude value is expected to be exceeded, on average, in each return period. If the yearly-maximum magnitude h is exceeded on average once in Tr-years, then its exceedance probability, $E(h) = 1 - G(h)$, in a given year is:

$$E(h) = P([H \geq h]) = \frac{1}{Tr(h)}$$

Therefore, the return period of the level value h is the inverse of the probability of exceedance and can be expressed as a function of the cumulative distribution, $G(h)$, of annual maxima, e.g. through the MEVD (Eq. 6):

$$Tr(h) = \frac{1}{E(h)} = \frac{1}{1 - G(h)} \tag{7}$$

Because for a fixed value of mean sea level there is a one-to-one relation between the value of the sum of the astronomical and the storm surge contribution, h, and the total ~~sea~~ water level, $z = h + msl$, one can write ~~$P(H > h) = P(H > z - msl) = P(Z > z = h + msl)$~~ $G_h(h) = P[H > h] = P[H > z - msl] = P[Z - msl > z - msl] = P[Z > z] = G_z(h)$, such that Eq. 7 can be used, once the cumulative distribution is known and for each (time-dependent) value of msl, to determine the return period of the total ~~sea level~~ water level (at the time when msl is evaluated):

$$Tr(z) = \frac{1}{1 - G_z(h)} = \frac{1}{1 - G_h(h)} = \frac{1}{1 - G(z - msl)} \tag{8}$$

Based on the hypothesis introduced in §2.2.6 that mean sea-level rise is the dominant effect in future coastal flooding, we assume that the characteristics of the extremes (i.e. the parameters of the GPDs defining the MEVD) remain valid in future scenarios. Eq. 8 clarifies that the return period of a fixed value z decreases as msl increases, basically because for higher values of msl a smaller value of y is needed to achieve the same total water level z. This decrease is non-linear, due to the nonlinear form of the right-hand side in Eq. 8.

3 Results and discussion

3.1 ~~Trend~~ Mann-Kendall trend analysis

We start by computing mean sea level on yearly basis and by subtracting it from observed total water level. The first question that we want explore is the presence of log-term trends~~in~~, unrelated to sea-level rise and associated to other factors (e.g., human-induced factors, morphological variations, etc.), in the "cleaned up" signal, i.e. the observed measurements without mean sea level. To answer this question, in this work we focus on the ~~observed time series. We compute the~~ deviation of yearly maxima from yearly mean sea level and test for the presence of trend by the two-tail Mann-Kendall test (Mann, 1945). Figure 2 summarizes results for each location explored. ~~We apply a linear fit to the yearly deviations from the yearly mean sea level and obtain: a) Venice time series~~ From a first visual inspection of Figure 2, the Venice (1872-2019) ~~have a slope of about 0.7 mm/year (p-value = 0.0141); b)~~ and Hornbæk ~~record~~ (1891-2012) ~~shows a trend of about 0.4 mm/year (p-value = 0.3525) ; c) Marseille sea levels~~ time series appear to show an increasing trend in the deviations of yearly maxima from yearly mean sea level (blue line) of different magnitudes. On the contrary, Marseille sea level observations (1985-2018) ~~have a linear trendwith slope of about -0.6 mm/year (p-value = 0.1109); d) Newlyn time series~~ seem to be characterized by a decreasing trend. Finally, the Newlyn historical record (1915-2016) ~~display a slope of 0.05 mm/year (p-value = 0.9976). The presence of a statistically significant trend is confirmed, with~~ displays a fairly constant signal with no noticeable variations. The application of the Mann-Kendall test reveals a partly different story. The test rejects the hypothesis of the absence of trend at the 95% confidence level, only for ~~Venice gauge station~~ the Venice site (p-value$^{\text{Venice}}$ = 0.014). This result suggests that the increase of the yearly maximum deviations from yearly mean sea level ~~is a direct result~~ may be a direct results of the local morphological variations of lagoon channels where the tidal wave propagates (whereby dissipation of the wave is reduced), and/or ~~vertical land motion (subsidence) that reduce the propagation and dissipation of the tide inside the Venice lagoon. Instead, for the other~~ land subsidence. On the contrary, at the remaining locations, the null hypothesis of no trend cannot ~~rejected through a Mann-Kendall test with 5% significance levels. In any case,~~ be rejected (p-value$^{\text{Hornbæk}}$ = 0.352, p-value$^{\text{Marseille}}$ = 0.110, and p-value$^{\text{Newlyn}}$ = 0.997). The results obtained from these analyses support the validity of the hypothesis that mean sea-level rise is the dominant factor in determining the future frequency of coastal flooding (see §2.2.6). For the tests performed here to compare different extreme-value statistical models, the ~~potentially confounding effect of the possible presence of trends is in our study avoided by the CV approach, which randomly reshuffles the sequence of observed years.~~ possible presence of trends (e.g. in Venice) is irrelevant, since such tests are performed by first reshuffling observed values, thereby eliminating any existing trend, albeit small.

3.2 Extreme value analysis

The MEVD formulation requires the choice of an optimal distribution of ordinary values that can represent the characteristics of the natural phenomenon under analysis. Different candidate distributions for the $F(x; \overrightarrow{\theta_j})$ in Eq. 6 are evaluated and the most suitable distribution is selected on the basis of the *CV* procedure comparing the MEVD-estimated quantiles with the observed ones. ~~According~~ As previously introduced in §2.2.3, according to different tests, the appropriate distribution to model
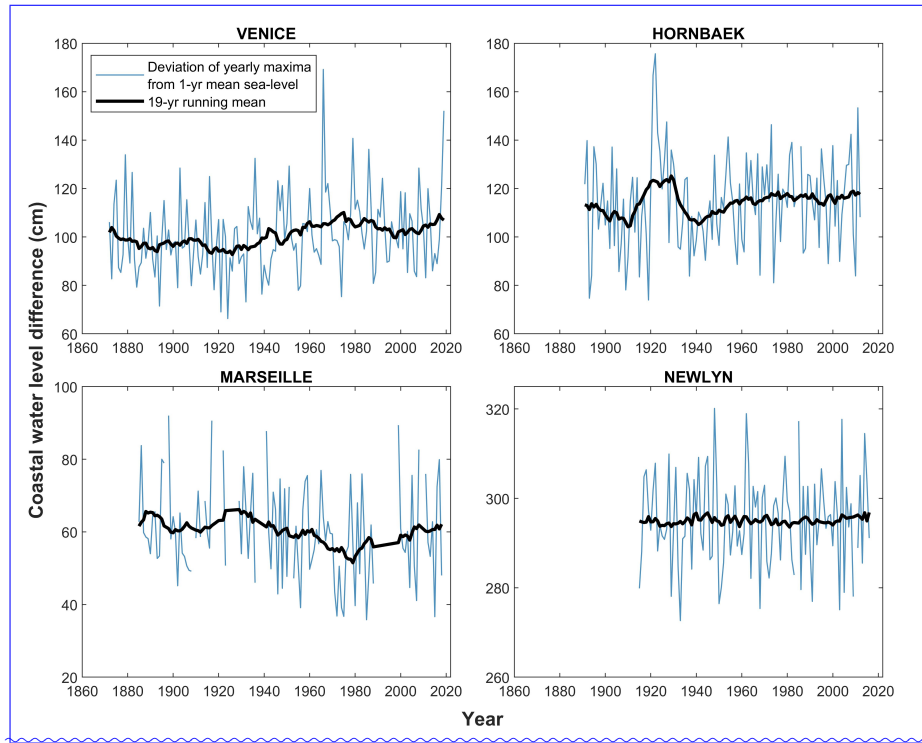
Figure 2. Deviation of yearly maxima from yearly mean sea level (blue line) ~~,~~ and 19-yr running mean (black line) calculated for Venice (IT), Hornbæk (DK), Marseille (FR), and ~~linear fit~~ Newlyn (~~red line~~UK)~~calculated for selected sites~~.

the ordinary sea-level values is the Generalized Pareto Distribution (GPD). We highlight that the GPD used in the MEVD framework is obtained by imposing a small threshold (differently from the high threshold adopted in the POT-GPD approach)
410 to capture the distribution of the main body of the probability distribution of the ordinary events and does require the event arrival process to be Poisson (Marani and Zorzetto, 2019).

As mentioned above (§2.2.4), the independence between two consecutive ~~storm~~ coastal water level events is guaranteed by imposing a minimum time lag. Firstly, we select daily maxima sea levels from the original record; secondly, define as independent events those that are separated by at least 30 days. Subsequently, the samples used for statistical inference are built as follows:
415 1) GEV-BM: the yearly maxima are selected; 2) ~~GEV-POT~~POT-GPD: as proposed by Coles (2001), the optimal threshold (u) is determined by studying the stability of the GPD shape (ξ) and modified scale ($\sigma^* = \sigma_u - \xi u$) parameters estimated using a wide range of values of u. Using this method, the following threshold values were identified: 65 cm (Venice), 50 cm (Hornbæk), 35 cm (Marseille) and 260 cm (Newlyn); 3) MEVD: all the independent ~~storm~~ coastal water level events above a low threshold are used to fit the probability distributions of ordinary values. ~~These thresholds are set to be large enough to filter out~~
420 ~~water level peaks that are likely to be associated to conditions without any storm contribution and sufficiently low to maximize~~ ~~the amount of information used. Based on local tidal ranges~~The optimal threshold to apply to all the independent events for extrapolating the ordinary values sample, is chosen by testing different threshold values and evaluating the goodness-of-fit of

the distribution using diagnostic graphical plots. According to the selection criteria described in §2.2.3, the low thresholds adopted at the four study sites are 59 cm for Venice, 40 cm for Hornbæk, 25 cm for Marseille, and 250 cm for Newlyn. For every observed site, Table 2 ~~displays~~ and Figure S3 display the gradual increase in the number of independent events (i.e. annual maxima, exceedances over the threshold, and ordinary values) used to infer the distributions when moving from GEV-BM, ~~GEV-POT~~ POT-GPD to MEVD approaches.

Considering the above threshold values, the observed and estimated distributions of ~~storm surge~~ coastal water level are compared by plotting their quantiles against each other. By comparing measures of in-sample and out-of sample test predictive accuracy, the results are presented by means of quantile-quantile (QQ) plots ~~in supplementary Figures S2, S3,~~ . The reader can refer to Figure 3 (or supplementary Figures S4, S5~~and~~, S6~~.~~, S7 and S8) to compare the results obtained with the MEVD framework (or the GEV-based approaches - GEV-BM and POT-GPD - vs. the MEVD formulation) for the four sites analyzed. QQ-plots are obtained as a result of the CV procedure with 1,000 random realizations and sample size: a) $S = 30$ years (in-sample-test on the left column); b) ~~$V = L - S$~~ $V = M - S$ years (out-of-sample test on the right column). The colours in ~~the~~ figures represent the density of points around the 45° line (i.e. the line of equality). This highlights how the estimated quantiles are closely comparable with the observed ones for all the three approaches tested and for both the sample size explored (S and V). In particular, ~~when~~ if the reader looks the supplementary Figures from S4 to S8 and if out-of-sample performance is considered, ~~MEVD~~ it is difficult to quantify which distribution is the best due to a large variability in the estimates. Overall, if only the MEVD performance is investigated, the reader can look the right column (out-of-sample test) in Figure 3, where the results display that the MEVD formulation performs similarly for all sites analyzed. In particular, it proves to be a good model for ~~the extreme sea levels.~~ lower/intermediate quantiles but shows variability in the estimates for higher quantiles.

Table 2. Total number of independent events and average number of events/year calculated for all observation sites.

Site name		Independent events		
		BM	POT	MEVD
Venice	Total	148	605	775
	N. events/year	1	4.08	5.23
Hornbæk	Total	121	595	736
	N. events/year	1	4.91	6.08
Marseille	Total	106	275	489
	N. events/year	1	2.57	4.61
Newlyn	Total	100	399	520
	N. events/year	1	3.99	5.20

We now focus on evaluating the performance of the three approaches (GEV-BM, ~~GEV-POT~~ POT-GPD and MEVD) in high-quantile estimation. We explore the predictive performance of the MEVD and GEV distribution as a function of the NDE (§2.2.5) computed for the maximum return period, $Tr_{max} = M - S + 1$, associated with the maximum value in each test sub-samples that we randomly extract in the CV approach. The use of NDE metric allows to easily characterize and compare models estimation uncertainty associated with fixed return time of interest and the variation of the calibration sample size (from 5 to 30 years). The results are summarized by means of box-plots (Figure ~~3~~4) and kernel density estimates computed for a calibration
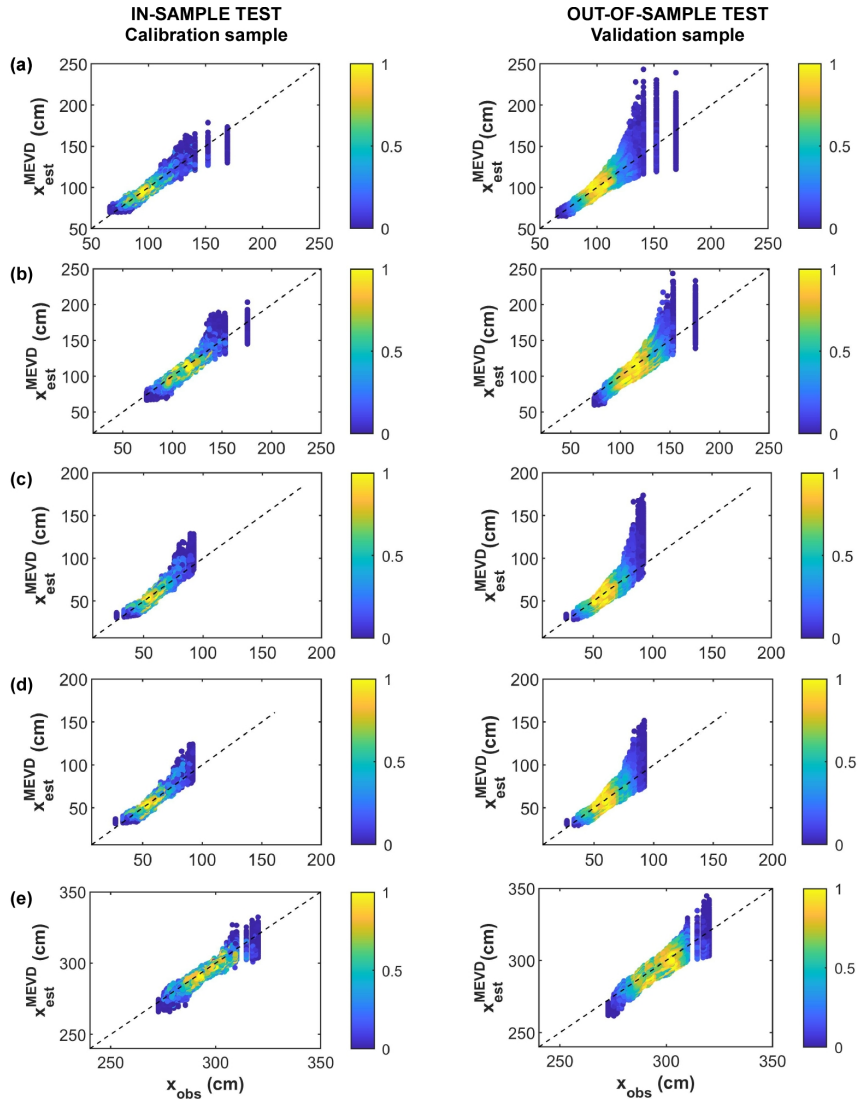
Figure 3. QQ-plots of extreme coastal water level quantiles, computed with the MEVD framework, for the (a) Venice (IT), (b) Hornbæk (DK), (c-d) Marseille (FR), and (e) Newlyn (UK) sites. The MEVD parameters estimations are based on non-overlapping sub-samples of fixed size (5 years), while subplots indicated with the letter 'd' display the QQ-plots obtained with MEVD parameters estimations based on data from the whole calibration sample size. The plots are obtained as a result of the cross validation method used to test the global performance of the models and are estimated for 1,000 random realizations and for sample size: a) S = 30 years (in-sample-test on the left column); b) V = M-S years (out-of-sample test on the right column). The colours represent the points density around the 45° line (black dashed line) corresponding to the best fit.

sample size of 30 years (Figure ~~4~~5). Table 3 summarizes the main results underlying the chosen evaluation metric. When we focus on the case of a short sample (5 years), different sites display variable results: I) the GEV and MEVD approaches perform

450 similarly for Venice (Figure ~~3~~4(a)) and Hornbæk (Figure ~~3~~4(b)) with similar interquartile ranges and underestimations of the actual quantile; II) for the Newlyn gauge station (Figure ~~3~~4(d)) the GEV-BM distribution yields better results, even though the POT and MEVD median error are also close to zero. On the contrary, when we consider longer calibration sample sizes (from 10 to 30 years), the MEVD-based estimates outperform the traditional approaches for most gauge stations explored: I) results for the Venice site confirm the robustness of the MEVD with respect to the GEV distribution especially for calibration sample

455 size equal to 30 years. In this case, the median error in the MEVD estimates tends to be closer to zero (-0.004), corresponding to approximately unbiased estimates; II) the Hornbæk station displays similar results to those for Venice and the MEVD-based estimates become more reliable when we consider a calibration sample size greater than 10-20 years; III) Newlyn estimation errors show a trade-off between the BM method and MEVD for calibration sample size equal to 20 and 30 years.

Results for the Marseille site show a peculiar behaviour that requires a specific discussion. In this case, the application of the

460 traditional extreme value theory is advantageous when compared with the MEVD (Figure ~~3~~4(c)). In order to better understand the application to the Marseille site, we performed MEVD parameter estimation using two approaches: 1) estimation based on non-overlapping calibration samples of fixed size (5 years as for the other sites); 2) parameter estimation on data from the whole calibration sample. The comparison of the results from these two set-ups confirms that when longer time slices are used for estimating GPD parameters (black colour in Figure ~~3~~4(c)), the MEVD performance is improved (for example

465 when we consider S = 30 years, MEVD median~~$[S-year_{window}]$~~$[S\text{-}year_{window}]$ = 0.17 vs. MEVD median~~$[5-year_{window}]$~~$[5\text{-}year_{window}]$ = 0.35), though it does not yet match the results obtained from GEV-BM approach (GEV-BM median error = 0.016). This can be explained by considering storm surges occur in Marseille about once every year on average. In this case GEV-BM is advantageous because the small number of events/year does not provide a more numerous calibration sample with respect to the sample of annual maxima. This result confirms the conclusion by Miniussi and Marani (2020), according to which the

470 selection of the estimation window size for fitting the ordinary value distribution strongly depends on the average number of extreme events/year.

We also provide a comparative analysis between the three methods to evaluate if the tested extreme value distributions are representative of the entire range of return times of interest. To achieve this purpose, we evaluate methods performance also for intermediate Tr values, greater than the calibration sample size, since for $Tr < S$ the empirical quantiles can be used.

475 We perform this additional analysis for the Venice, Hornbæk and Newlyn sites. Figure 6 summarizes the results obtained by estimating the probability distribution parameters on 30-year calibration sub-samples. The analyses suggest that when we focus on the median error associated with moderate values of the return period, GEV-BM displays an overall greater robustness (e.g., in the case of Venice and Hornbæk sites) with respect to POT-GPD and MEVD, which exhibit greater fluctuations. In particular, results show that MEVD is a good model for the highest values of the return period, but exhibit a greater absolute

480 value of the estimation error for smaller Tr. Overall, the results suggest that no single approach is clearly superior at all values of Tr, due to a large variability in the estimates. For example, for the Venice site there is a decrease (in many cases an unbiased estimates) in the MEVD-NDE values for intermediate Tr (between 85 and 105 years) while for greater Tr values (but smaller

Table 3.

Results of the evaluation metric obtained for all the gauge stations and for calibration sample sizes (S) equal to 5 and 30 years. In the case of the Marseille site, text in bold refers to MEVD parameter estimation based on data from the whole calibration sample size.

Site name	Variables	S = 5 yrs			S = 30 yrs		
		BM	POT	MEVD	BM	POT	MEVD
Venice	NDE-median	-0.160	-0.175	-0.178	-0.133	-0.158	-0.004
	NDE-mean	-0.069	-0.101	-0.116	-0.087	-0.133	~~-0.024~~0.024
	NDE-std	0.366	0.274	0.267	0.156	0.113	0.155
Hornbæk	NDE-median	-0.119	-0.104	-0.113	-0.113	-0.115	0.056
	NDE-mean	-0.069	-0.101	-0.116	-0.068	-0.087	0.077
	NDE-std	0.366	0.274	0.267	0.113	0.100	0.131
Marseille	NDE-median	-0.0003	0.059	0.172	0.016	0.047	0.357 **0.172**
	NDE-mean	0.045	0.129	0.262	0.013	0.050	0.374 **0.183**
	NDE-std	0.252	0.350	0.421	0.072	0.115	0.178 **0.140**
Newlyn	NDE-median	-0.010	-0.030	-0.033	~~-0.033~~-0.003	-0.032	0.0008
	NDE-mean	0.003	-0.022	-0.026	-0.002	-0.031	~~0.0020~~0.002
	NDE-std	0.050	0.042	0.042	0.016	0.014	~~0.0210~~0.021

than Tr_{max}) the error shows an overestimation of the actual quantile with respect to traditional approaches (which exhibit an underestimation tendency). To be more specific, if $Tr > 105$ years are considered, MEVD yields error estimates between zero and <10%, while errors associated with GEV-BM and POT-GPD lie between zero and <-20%. The Hornbæk site shows similar results to the Venice site, while Newlyn's results exhibit more fluctuations for large Tr values with much reduced smaller amplitudes and values of the NDE.

485
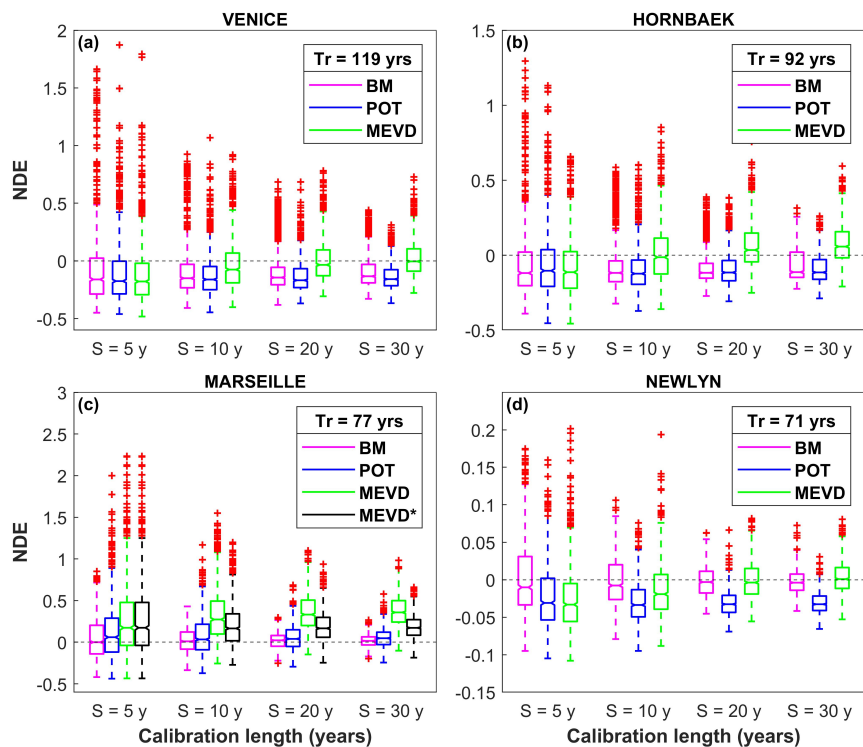
Figure 4. Distribution of the Non Dimensional Error (NDE) for maximum sample return period (Tr) represented by means of box-plots at given gauge stations explored: (a) Venice (IT), (b) Hornbæk (DK), (c) Marseille (FR), (d) Newlyn. In the case of the Marseille (FR) site, MEVD parameter estimation is based: 1) green colour: on non-overlapping sub-samples of fixed size (5 years), and 2) black colour: on data from the whole calibration sample.
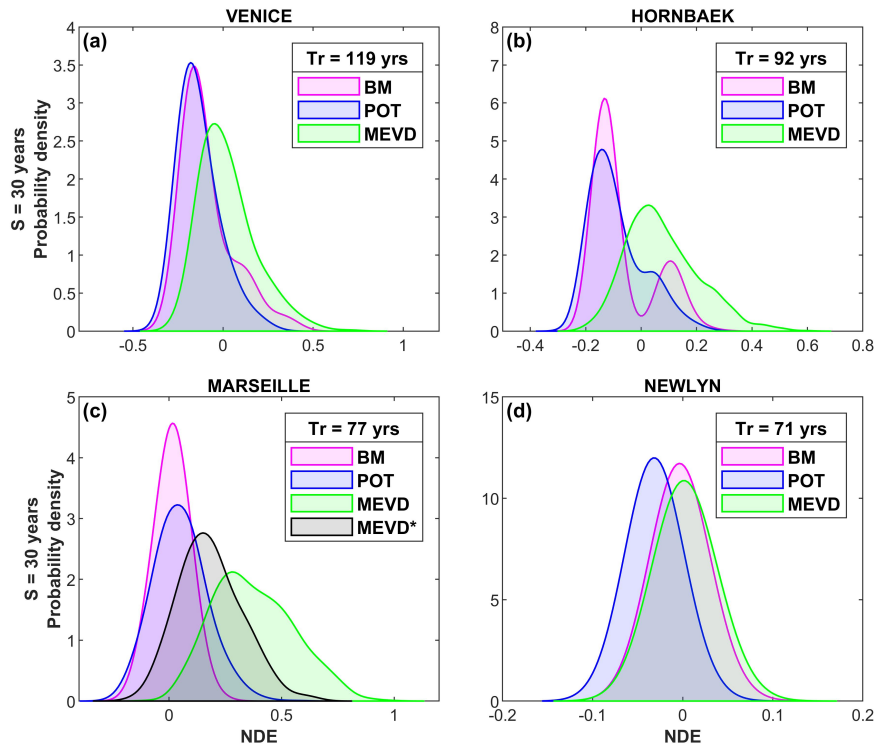
Figure 5. Kernel density estimates for the Non Dimensional Error (NDE) distributions obtained with a calibration sample size (S) of 30 years and maximum return period (Tr) at given gauge stations explored: (a) Venice (IT), (b) Hornbæk (DK), (c) Marseille (FR), (d) Newlyn (UK). In the case of the Marseille (FR) site, MEVD parameter estimation is based: 1) green ~~coluor~~colour: on non-overlapping sub-samples of fixed size (5 years), and 2) black colour: on data from the whole calibration sample.
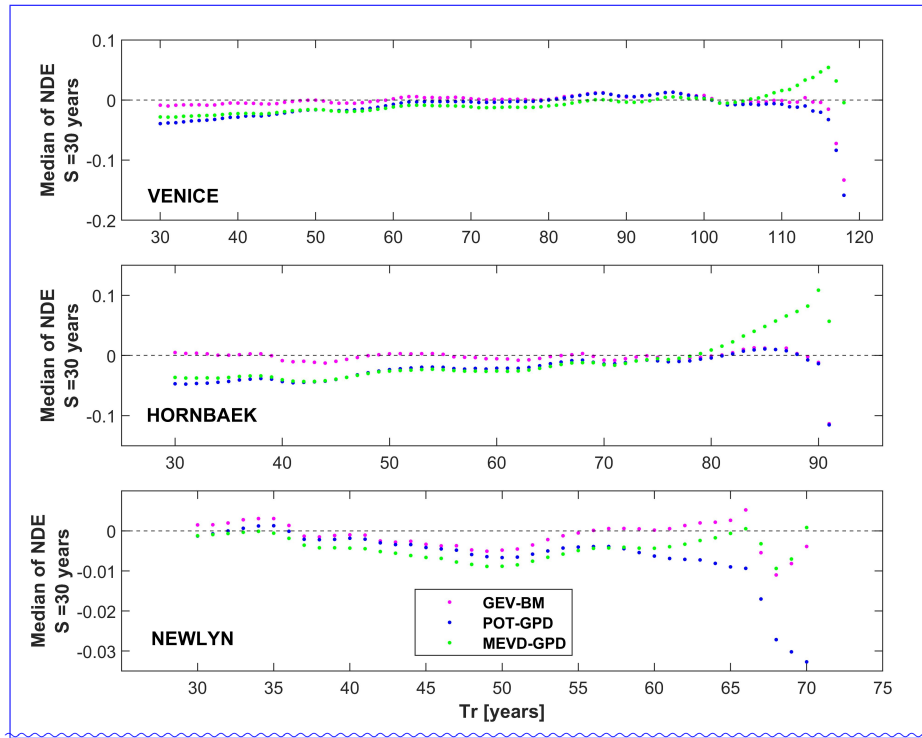
Figure 6. Median of the non-dimensional error (NDE) for return period greater than the calibration sample size in test sub-sample for the GEV-BM, POT-GPD and MEVD approaches (magenta, blue and green dots respectively). The results are obtained for the Venice (IT), Hornbæk (DK) and Newlyn (UK) sites and by estimating the distribution parameters on 30-year calibration sub-samples.

3.3 Future projections of extreme storm surges

490 We next explore how sea-level rise may influence the frequency of extreme ~~storm surges~~ total water levels across the sites analyzed. As described in §2.2.6, we only evaluate the influence of an increased mean sea level, i.e. we do not address possible changes in storm regimes (e.g., see Tebaldi et al., 2012).

We use site-specific sea-level projections from IPCC's AR5 (Church et al., 2013), which indicate an accelerating sea-level rise at all four observation sites (for each gauge station under analysis, the reader can refer to the panels (a), (c), (e), and (g) in

495 Figure ~~5~~7), with expected water level increases at the end of the century (RCP8.5) of 48 cm in Venice, 52 cm in Hornbæk, 59 cm in Newlyn and 54 cm in Marseille. The panels (b), (d), (f), and (h) in Figure ~~5~~7 show observed (green line) and future (blue and red lines) changes in the return period associated with event maximum water levels due to sea-level rise. These curves were obtained by using, in Eq. 8, the MEVD with parameters estimated on 5-year sliding windows. As noted above, changes in return levels are nonlinear way: relative changes are more significant for smaller extremes than for larger ones. The Tr

500 vs. z curves are concave downward and display varying slopes depending on the site explored. When a fixed return period is considered (e.g., 500 years), the mean sea level projections quantify the expected increase in extreme ~~storm-surge~~ coastal water

Table 4.

Results of the percentage changes in water level (Δz) obtained with the two future scenarios (RCP4.5 and RCP8.5) and the return times (100 and 500 years) at the four sites under analysis.

Tr (yrs)	RCP	Δz (%)			
		Venice	Hornbæk	Marseille	Newlyn
100	4.5	16.75%	14.22%	22.62%	11.29%
	8.5	22.82%	21.76%	29.73%	15.26%
500	4.5	14.60%	11.70%	16.24%	11.23%
	8.5	19.92%	18.18%	21.91%	15.09%

level peaks for that particular return period. These changes vary heterogeneously across the different coastal sites explored. By comparing the percentage changes associated with the two emission scenarios and the two return periods (Table 4), Venice and Marseille are seen to experience the greatest changes in extreme water levels (e.g., with reference to $Tr = 100$ years and RCP8.5, the variations at Venice and Marseille sites are approximately 23% and 39% respectively). All sites display greater percent changes for the lower 100-year return period in each scenario, i.e. "less-infrequent" extremes will be most impacted by sea-level changes in the near future.

Changes in sea-level extremes can also be studied by focusing on changes in the return period of a fixed value of the water level. To this end, one can define a sensitivity measure as:

$$SM = \frac{1}{Tr} \cdot \frac{dTr}{dmsl} = -\frac{1}{Tr} \cdot \frac{1}{[1 - G(z - msl)]^2} \cdot f(z - msl) = -f(z - msl) \cdot Tr \qquad (9)$$

which is obtained by derivation of Eq. 8, and where $f(z) = \frac{dG}{dz}$ is the probability density function associated with $G(z)$. Eq. 9 shows that, at a given site and for a set value of z, the relative change in return period grows linearly with Tr. For example, see in Figures 56b, d, f, h how, for a given value of z, changes (horizontal spacing between the curves) are greater for $Tr = 1000$ years than for $Tr = 500$ years. The expression for SM also tells us that changes in Tr are more significant, everything else being equal, for values of $z - msl$ near the mode of the distribution, where $f(z - msl)$ is maximum (e.g., compare changes at the Venice or Hornbæk sites with those at Newlyn for a same initial value of Tr).

4 Conclusions

~~In this study, we apply for the first time the Metastatistical Extreme Value Distribution to extreme sea levels observed~~ The comparative examination of extreme value distributions applied to observed sea levels at several sites along European coasts provides insights into the predictive performance of traditional and new approaches. Our analyses confirm some practical and conceptual advantages of the MEVD with respect to traditional methods ~~used for studying extreme events.~~ A cross-validation scheme (with 1,000 realizations for each sites) ~~is~~ was used to compare model ~~performances~~ performance in high-quantile estimation. The use of two independent sub-sample (calibration and test sample) allows ~~to quantify the "real" predictive uncertaintyin the high-return period quantiles estimation.In this manner, it is possible to evaluate the capacity of the models to~~
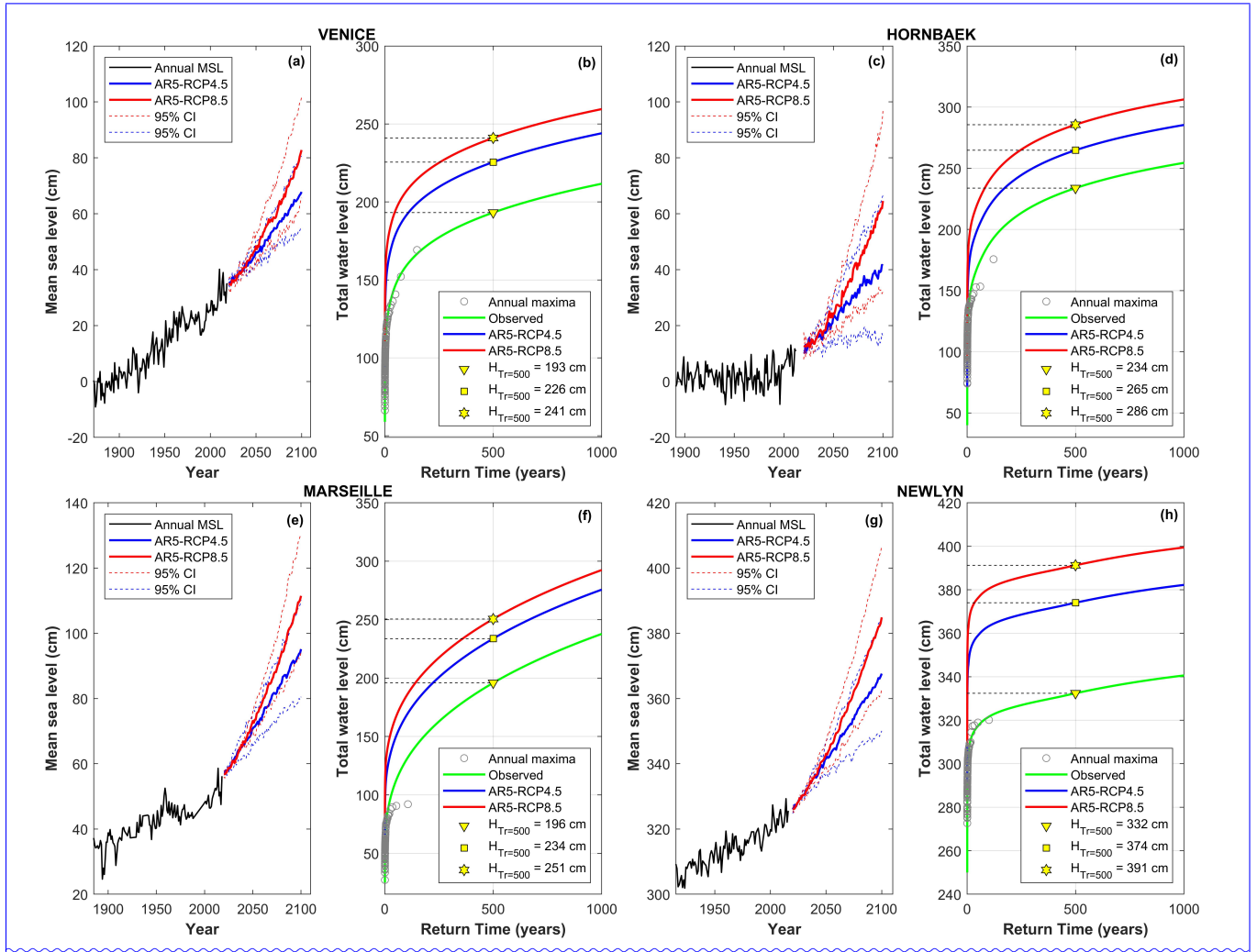
Figure 7. Future ~~storm surge~~ total water level projections at: Venice (IT, panels (a) and (b)), Hornbæk (DK, panels (c) and (d)), Marseille (FR, panels (e) and (f)), and Newlyn (UK, panels (g) and (h)). The panels represent: 1) (a), (c), (e), and (g): annual (black line) and future mean sea level until 2100 with RCP4.5 (blue line) and RCP8.5 (red line). Dashed line represents the 95% confidence intervals. 2) (b), (d), (f), and (h): return period curves for extreme total water level. The green curve represents the estimates obtained with the observed record; the blue and red curves represent the estimates obtained with the projected sea-level rise (SLR) until 2100 with RCP4.5 (blue) and RCP8.5 (red) respectively; the grey dots indicate the observed annual maxima. Triangle, square and pentagon highlight the heights of extreme storm surge for a fixed return period equal to 500 years.

~~extrapolate from observed extremes and to predict the frequency of occurrence of unobserved ones.~~the quantification of actual predictive uncertainty.

~~To address the problems posed by observational samples of different sizes, we explored the effect on uncertainty of different calibration sample sizes, from 5 to 30 years.~~ We find that the MEVD approach provides reliable estimates of high quantiles ~~in~~ for almost all the gauge stations explored, ~~especially when we consider longer~~ particularly when sufficiently long calibration sample sizes ~~.~~are considered. Differences in performance between the MEVD framework and GEV-based approaches are not large, and a definitive conclusion on an optimal solution independent of the return period of interest remains elusive. However, small differences in the estimation accuracy are relevant for engineering applications when dealing with rare extreme events. If we focus on high-return period quantiles estimation, our analyses show that the MEVD approach provides reliable estimates for almost all the gauge stations explored. Data from the Marseille gauge station exhibit a behaviour that deviates from those from other sites, showing an inferior predictive performance of the MEVD with respect to GEV-based approaches. We interpret this fact to be linked to ~~other factors (i.e. morphology, topography, bathymetry, tidal driven and atmosphere forcing), which result in a small number of large~~ the small average number of sea-level peaks every year~~. The small number~~: the small sample of yearly ordinary events ~~is such that the MEVD can use little additional~~ available prevents the MEVD from adding significant information with respect to GEV-BM and ~~GEV-POT.~~ POT-GPD. Conversely, when we evaluate methods performance for intermediate return period values, GEV-BM displays an overall greater robustness, and MEVD exhibit a greater absolute value of the estimation error.

~~Finally, we explore the variations of the frequency of extreme storm water levels under future sea-level scenarios.~~Analyses using sea-level projections through 2100 quantify the increase in extreme storm water levels associated with fixed Unfortunately, the size of the available datasets does not allow to explore model performance for greater values of the return period. ~~Conversely,~~ Future work could investigate if the estimation error can be reduced, with respect to what was found here, by using different approaches, e.g., by assuming "time-invariant" parameters in the ordinary distribution, whose estimation would thus be performed on the entire calibration dataset, rather than on relatively short sliding windows. Synthetic water level time series may be produced by one of the several existing numerical models to extend analyses to arbitrarily long return period values.

Finally, we explored projections of the ~~return period of a fixed value of the water level decreases with increasing mean sea levels~~frequency of extreme total water levels driven by changes in mean sea level. The sensitivity of extreme water level frequency to sea-level rise is location-dependent ~~, and the introduced sensitivity measure shows~~ and we find that, at a given site and for a set value of the ~~water level~~total water level extreme, the relative change in return time grows linearly with ~~the value of the latter. The MEVD methodology proposed in this work, by relaxing some of the hypotheses underlying the traditional extreme value distribution, reduces the uncertainty associated with the assessment of coastal flooding risk under changing sea-levels.~~ return time itself.

Data availability. All data used are publicly available from sources cited in the main text. Venice sea-level data were obtained from the "Centro Previsioni e Segnalazioni Maree" of the Venice Municipality (https://www.comune.venezia.it/it/content/centro-previsioni-e-segnalazioni-maree). The remaining water level data were downloaded from the University of Hawaii Sea Level Center (UHSLC) repository (http://uhslc.soest.hawaii.edu/data/?rq#uh745a/). The CMIP5 model outputs used for the future total water level projections are available at the "Integrated Climate Data Center" (ICDC) database of the University of Hamburg (https://icdc.cen.uni-hamburg.de/en/ar5-slr.html).

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. M.F.C. acknowledges the PhD School in Civil Sciences and Environmental Engineering by University of Padova for funding her PhD. M.M. acknowledges financial support by the "Scientific activity performed with the contribution of the Provveditorato for the Public Works of Veneto, Trentino Alto Adige, and Friuli Venezia Giulia, provided through the concessionary of State Consorzio Venezia Nuova and coordinated by CORILA in the framework of the Venezia 2021 Research Program".

References

Araújo, I. B. and Pugh, D. T.: Sea levels at Newlyn, 1915–2005: Analysis of trends for future flooding risks, J. Coastal Res., 24, 03–212, https://doi.org/10.2112/06-0785.1, 2008.

Balkema, A. A. and de Haan, L.: Residual life time at great age, Ann. Probab., 2, 792–804, https://doi.org/10.1214/aop/1176996548, 1974.

Barbariol, F., Bidlot, J.-R., Cavaleri, L., Sclavo, M., Thomson, J., and Benetazzo, A.: Maximum wave heights from global model reanalysis, Prog. Oceanogr., 175, 139–160, https://doi.org/10.1016/j.pocean.2019.03.009, 2019.

Beck, C. and Cohen, E. G. D.: Superstatistics, Physica A: Statistical Mechanics and its Applications, 322, 267–275, https://doi.org/10.1016/S0378-4371(03)00019-0, 2003.

Beirlant, J., Goegebeur, Y., Segers, J. J. J., and Teugels, J.: Statistics of Extremes: Theory and Applications, John Wiley & Sons, Chichester, UK, 2004.

Benetazzo, A., Ardhuin, F., Bergamasco, F., Cavaleri, L., Guimarães, P. V., Schwendeman, M., Sclavo, M., Thomson, J., and Torsello, A.: On the shape and likelihood of oceanic rogue waves, Sci. Rep, 7, 1–11, https://doi.org/10.1038/s41598-017-07704-9, 2017.

Bernardara, P., Andreewsky, M., and Benoit, M.: Application of regional frequency analysis to the estimation of extreme storm surges, J. Geophys. Res.-Oceans, 116, https://doi.org/10.1029/2010JC006229, 2011.

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

Bernier, N. B. and Thompson, K. R.: Tide-surge interaction off the east coast of Canada and northeastern United States, J. Geophys. Res.-Oceans, 112, https://doi.org/10.1029/2006JC003793, 2007.

Bommier, E.: Peaks-over-threshold modelling of environmental data, Master's thesis, Department of Mathematics, Uppsala University, 2014.

Cancelliere, A.: Non Stationary Analysis of Extreme Events, Water Resources Management, 31, 3097–3110, https://doi.org/10.1007/s11269-017-1724-4, 2017.

Castillo, E., Hadi, A. S., Balakrishnan, N., and Sarabia, J. M.: Extreme Value and Related Models in Engineering and Science Applications, John Wiley & Sons, New York, 2005.

Chan, S., Chu, J., Zhang, Y., and Nadarajah, S.: An extreme value analysis of the tail relationships between returns and volumes for high frequency cryptocurrencies, Research in International Business and Finance, 59, 101 541, https://doi.org/10.1016/j.ribaf.2021.101541, 2022.

Chiu, Y., Chebana, F., Abdous, B., Bélanger, D., and Gosselin, P.: Mortality and morbidity peaks modeling: an extreme value theory approach, Stat. Methods Med. Res., 27, 1498–1512, https://doi.org/10.1177/0962280216662494, 2018.

Chow, V. T.: Frequency analysis of hydrologic data with special application to rainfall intensities, University of Illinois Engineering Experiment Station, Bulletin 414, 1953.

Church, J. A. and White, N. J.: A 20th century acceleration in global sea-level rise, Surv. Geophys., 33, https://doi.org/10.1029/2005GL024826, 2006.

Church, J. A. and White, N. J.: Sea-Level Rise from the Late 19th to the Early 21st Century, Surv. Geophys., 32, https://doi.org/10.1007/s10712-011-9119-1, 2011.

Church, J. A., Clark, P. U., Cazenave, A., Gregory, J. M., Jevrejeva, S., Levermann, A., Merrifield, M. A., Milne, G. A., Nerem, R. S., Nunn, P. D., Payne, A. J., Pfeffer, W. T., Stammer, D., and Unnikrishnan, A. S.: Sea Level Change, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited

by Stocker, T., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., chap. 13, pp. 1137–1216, Cambridge University Press, https://www.ipcc.ch/report/ar5/wg1/, 2013.

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. Dynam., 46, 1503–1516, https://doi.org/10.1007/s00382-015-2659-1, 2015.

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

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.

Dalrymple, T.: Flood frequency Analysis, Manual of hydrology: Part 3. Flood-flow techniques 1543, Geological Survey Water Supply Paper 1543-A, https://doi.org/10.3133/wsp1543A, 1960.

Davison, A. C. and Smith, R. L.: Models for Exceedances over High Thresholds, J. Roy. Stat. Soc. B-Met., 52, 393–442, http://www.jstor.org/stable/2345667, 1990.

De Zea Bermudez, P. and Mendes, Z.: Extreme value theory in medical sciences: modeling total high cholesterol levels, Journal of Statistical Theory and Practice volume, 6, 468–491, https://doi.org/10.1080/15598608.2012.695673, 2012.

Dixon, M. J. and Tawn, J. A.: The effect of non-stationarity on extreme sea-level estimation, J. Roy. Stat. Soc. C-App., 48, 135–151, 1999.

Dubey, S. D.: A compound weibull distribution, Naval Research Logistics Quarterly, 15, 179–188, https://doi.org/10.1002/nav.3800150205, 1968.

Eliot, M.: Influence of interannual tidal modulation on coastal flooding along the Western Australian coast, J. Geophys. Res.-Oceans, 115, https://doi.org/10.1029/2010JC006306, 2010.

Elvidge, S. and Angling, M. J.: Using Extreme Value Theory for Determining the Probability of Carrington-Like Solar Flares, Space Weather, 16, 417–421, https://doi.org/10.1002/2017SW001727, 2018.

Embrechts, P., Klüppelberg, C., and Mikosch, T.: Modelling Extremal Events for insurance and finance, Springer, New York, 1997.

Ferro, C. A. T. and Segers, J.: Inference for clusters of extreme values, J. Roy Stat. Soc. B-Met., 65, 545–556, https://doi.org/10.1111/1467-9868.00401, 2003.

Finkenstädt, B. and Rootzén, H.: Extreme Values in Finance, Telecommunications and the Environment, Taylor & Francis Inc, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Boca Raton, Florida, 2004.

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.

Fortunato, A. B., Li, K., Bertin, X., Rodrigues, M., and Miguez, B. M.: Determination of extreme sea levels along the Iberian Atlantic coast, Ocean Eng., 111, 471–482, https://doi.org/https://doi.org/10.1016/j.oceaneng.2015.11.031, 2016.

Fréchet, M. R.: Sur la loi de probabilité de l'écart maximum, Ann. Soc. Polon. Math., 6, 93–116, 1927.

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

Goring, D. G., Stephens, S. A., Bell, R. G., and P., P. C.: Estimation of Extreme Sea Levels in a Tide-Dominated Environment Using Short Data Records, Journal of Waterway, Port, Coastal, and Ocean Engineering, 137, 150–159, https://doi.org/10.1061/(ASCE)WW.1943-5460.0000071, 2011.

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.

Gumbel, E. J.: Les valeurs extrêmes des distributions statistiques, Annales de l'Institut Henri Poincaré, 5, 115–158, www.numdam.org/item/AIHP_1935__5_2_115_0/, 1935.

Gumbel, E. J.: Statistics of Extremes, Columbia University Press, New York, 1958.

Haigh, I. D., 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.

Haigh, I. D., Eliot, M., and Pattiaratchi, C.: Global influences of the 18.61-year nodal cycle and 8.85-year cycle of lunar perigee on high tidal levels, J. Geophys. Res.-Oceans, 116, 25 249, https://doi.org/10.1029/2010JC006645, 2011.

Haigh, I. D., MacPherson, L. R., Mason, M. S., Wijeratne, E. M. S., Pattiaratchi, C. B., Crompton, R. P., and George, S.: Estimating present day extreme water level exceedance probabilities around the coastline of Australia: tropical cyclone-induced storm surges, Clim. Dynam., 42, 139–157, https://doi.org/10.1007/s00382-012-1653-0, 2014a.

Haigh, I. D., Wahl, T., Rohling, E. J., Price, R. M., Pattiaratchi, C., Calafat, F. M., and Dangendorf, S.: Timescales for detecting a significant acceleration in sea level rise, Nat. Commun., 5, https://doi.org/10.1038/ncomms4635, 2014b.

Hall, T. M. and Sobel, A. H.: On the impact angle of Hurricane Sandy's New Jersey landfall, Geophys. Res. Lett., 40, 2312–2315, https://doi.org/10.1002/grl.50395, 2013.

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

Hamdi, Y., Bardet, L., Duluc, C.-M., and Rebour, V.: Use of historical information in extreme-surge frequency estimation: the case of marine flooding on the La Rochelle site in France, Nat. Hazard Earth Sys., 15, 1515–1531, https://doi.org/10.5194/nhess-15-1515-2015, 2015.

Hamdi, Y., Garnier, E., Giloy, N., Duluc, C.-M., and Rebour, V.: Analysis of the risk associated with coastal flooding hazards: a new historical extreme storm surges dataset for Dunkirk, France, Nat. Hazard Earth Sys., 18, 3383–3402, https://doi.org/10.5194/nhess-18-3383-2018, 2018.

Hamon, B. V. and Middleton, J. F.: Return periods of extreme sea levels: the exceedance probability method, International Hydrographic Review, 66, 165–177, 1989.

Hay, C., Morrow, E., Kopp, R., and Mitrovica, J. X.: Probabilistic reanalysis of twentieth-century sea-level rise, Nature, 517, 481–484, https://doi.org/10.1038/nature14093, 2015.

Hosking, J. R. M.: L-moments: Analysis and estimation of distributions using linear combinations of order statistics, J. Roy. Stat. Soc. B-Met., 52, 105–124, 1990.

Hosking, J. R. M. and Wallis, J. R.: Parameter and Quantile Estimation for the Generalized Pareto Distribution, Technometrics, 29, 339–349, http://www.jstor.org/stable/1269343, 1987.

Hosking, J. R. M. and Wallis, J. R.: Some statistics useful in regional frequency analysis, Water Resour. Res., 29, 271–281, https://doi.org/10.1029/92WR01980, 1993.

Hosking, J. R. M. and Wallis, J. R.: Regional frequency analysis: an approach based on L-moments, Cambridge University press, https://doi.org/10.1017/CBO9780511529443, 1997.

Hosking, J. R. M., Wallis, J. R., and Wood, E. F.: Estimation of the generalized extreme-value distribution by the method of probability-weighted moments, Technometrics, 27, 251–261, https://doi.org/10.1080/00401706.1985.10488049, 1985.

Hosseini, S. R., Scaioni, M., and Marani, M.: Extreme Atlantic hurricane probability of occurrence through the Metastatistical Extreme Value Distribution, Geophys. Res. Lett., 47, 2019GL086 138, https://doi.org/10.1029/2019GL086138, 2020.

Jenkinson, A.: The frequency distribution of the annual maximum (or minimum) values of meteorological elements, Q. J. Roy. Meteor. Soc., 81, 158–171, 1955.

680 Jevrejeva, S., Moore, J. C., Grinsted, A., and Woodworth, P. L.: Recent global sea level acceleration started over 200 years ago?, Geophys. Res. Lett., 35, https://doi.org/10.1029/2008GL033611, 2008.

Johns, B. and Ali, M. A.: The numerical modeling of storm surges in the Bay of Bengal, Q. J. Roy. Meteor. Soc., 106, 1–18, https://doi.org/10.1002/qj.49710644702, 1980.

Katz, R. W., Parlange, M. B., and Naveau, P.: Statistics of extremes in hydrology, Adv. Water Resour., 25, 1287–1304, 685 https://doi.org/10.1016/S0309-1708(02)00056-8, 2002.

Katz, R. W., Brush, G. S., and Parlange, M. B.: Statistics of extremes: modeling ecological disturbances, Ecology, Ecology, 86, 1124–1134, https://doi.org/10.1890/04-0606, 2005.

Li, G., Zhang, X., Zwiers, F., and Wen, Q. H.: Quantification of uncertainty in high-resolution temperature scenarios for North America, J. Climate, 25, 3373–3389, https://doi.org/10.1175/JCLI-D-11-00217.1, 2012.

690 Lionello, P., Sanna, A., Elvini, E., and Mufato, R.: A data assimilation procedure for operational prediction of storm surge in the northern Adriatic Sea, Cont. Shelf. Res., 26, 539–5539, https://doi.org/10.1016/j.csr.2006.01.003, 2006.

Lowe, J. A., Woodworth, P. L., Knutson, T., McDonald, R. E., McInnes, K., Woth, K., Von Storch, H., Wolf, J., Swail, V., Bernier, N., Gulev, S., Horsburgh, K., Unnikrishnan, A. S., Hunter, J., and Weisse, R.: Past and future changes in extreme sea levels and waves, in: Understanding Sea-Level Rise and Variability, edited by Church, J. A., Woodworth, P. L., Aarup, T., and Wilson, W. S., chap. 11, pp. 695 326–375, Wiley-Blackwell, https://doi.org/10.1002/9781444323276, 2010.

Mann, H. B.: Nonparametric Tests Against Trend, Econometrica, 13, 245–259, 1945.

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.

Marani, M. and Zorzetto, E.: Doubly stochastic distributions of extreme events, arXiv, https://arxiv.org/abs/1902.09862, 2019.

700 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/https://doi.org/10.1016/j.advwatres.2018.05.001, 2018.

McInnes, K. L., Hubbert, G., Macadam, I., and O'Grady, J.: An assessment of current and future vulnerability to coastal inundation due to sea-level extremes in Victoria, southeast Australia, Int. J. Climatol., 33, 33–47, https://doi.org/10.1002/JOC.3405, 2013.

705 Meehl, G., Stocker, T., Collins, W. D., Friedlingstein, P., Gaye, A., Gregory, J. M., Kitoh, A., Knutti, R., Murphy, J. M., Noda, A., Raper, S. C. B., Watterson, I. G., Weaver, A. J., and Zhao, Z. C.: Global Climate Projections, in: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., Tignor, M., and Miller, H., chap. 10, pp. 747– 846, Cambridge University Press, http://www.ipcc.ch/publications_and_data/publications_ipcc_fourth_assessment_report_wg1_report_ 710 the_physical_science_basis.htm, 2007.

Mekonnen, K., Melesse, A. M., and Woldesenbet, T. A.: Effect of temporal sampling mismatches between satellite rainfall estimates and rain gauge observations on modelling extreme rainfall in the Upper Awash Basin, Ethiopia, Journal of Hydrology, 598, 126 467, https://doi.org/10.1016/j.jhydrol.2021.126467, 2021.

Menéndez, M. and Woodworth, P. L.: Changes in extreme high water levels based on a quasi-global tide-gauge data set, J. Geophys. Res., 715 115, 1124–1134, https://doi.org/10.1029/2009JC005997, 2010.

Middleton, J. F. and Thompson, K. R.: Return periods of extreme sea levels from short records, Journal of Geophysical Research, 91, 11 707–11 716, https://doi.org/10.1029/JC091iC10p11707, 1986.

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, e2019WR026 535, https://doi.org/0.1029/2019WR026535, 2020.

Miniussi, A. and Marra, F.: Estimation of extreme daily precipitation return levels at-site and in ungauged locations using the simplified MEV approach, Journal of Hydrology, 603, 126 946, https://doi.org/10.1016/j.jhydrol.2021.126946, 2021.

Miniussi, A., Marani, M., and Villarini, G.: Metastatistical Extreme Value Distribution applied to floods across the continental United States, Adv. Water Resour., 136, 103 498, https://doi.org/10.1016/j.advwatres.2019.103498, 2020a.

Miniussi, A., Villarini, G., and Marani, M.: Analyses through the metastatistical extreme value distribution identify contributions of tropical cyclones to rainfall extremes in the eastern United States, Geophys. Res. Lett., 47, e2020GL087 238, https://doi.org/10.1029/2020GL087238, 2020b.

Önöz, B. and Bayazit, M.: Effect of the occurrence process of the peaks over threshold on the flood estimates, J. Hydrol., 244, 86–96, https://doi.org/10.1016/S0022-1694(01)00330-4, 2001.

Oppenheimer, M., Glavovic, B., Hinkel, J., van de Wal, R., Magnan, A., Abd-Elgawad, A., Cai, R., Cifuentes-Jara, M., DeConto, R., Ghosh, T., Hay, J., Isla, F., Marzeion, B., Meyssignac, B., and Sebesvari, Z.: Sea Level Rise and Implications for Low-Lying Islands, Coasts and Communities, in: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate, edited by Pörtner, H.-O., Roberts, D., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegría, A., Nicolai, M., Okem, A., Petzold, J., Rama, B., and Weyer, N. M., chap. 4, pp. 321–445, Cambridge University Press, https://www.ipcc.ch/srocc/download-report/, 2019.

Papalexiou, S. M. and Koutsoyiannis, D.: Battle of extreme value distributions: A global survey on extreme daily rainfall, Water Resources Research, 49, 187–201, https://doi.org/10.1029/2012WR012557, 2013.

Peng, D., Hill, E. M., and Meltzner, A. J.and Switzer, A. D.: Tide gauge records show that the 18.61-year nodal tidal cycle can change high water levels by up to 30 cm, J. Geophys. Res.-Oceans, 124, 736–749, https://doi.org/10.1029/2018JC014695, 2019.

Pickands, III, J.: Statistical inference using extreme order statistics, Ann. Stat., 3, 119–131, https://doi.org/10.1214/aos/1176343003, 1975.

Pisarenko, V. F., Rodkin, M. V., and Rukavishnikova, T. A.: Estimation of the probability of strongest seismic disasters based on the extreme value theory, Izvestiya, Physics of the Solid Earth, 50, 311–324, https://doi.org/10.1134/S1069351314030070, 2014a.

Pisarenko, V. F., Sornette, A., Sornette, D., and Rodkin, M. V.: Characterization of the Tail of the Distribution of Earthquake Magnitudes by Combining the GEV and GPD Descriptions of Extreme Value Theory, Pure Appl. Geophys., 171, 1599–1624, https://doi.org/10.1007/s00024-014-0882-z, 2014b.

Pugh, D. T. and Vassie, J. M.: Extreme sea levels from tide and surge probability, in: Proc. 16th International Conference on Coastal Engineering 1978, Hamburg, Germany, chap. 52, pp. 911–930, American Society of Civil Engineers, 1979.

Pugh, D. T. and Vassie, J. M.: Applications of the joint probability method for extreme sea level computations, in: Proc. Inst. Civ. Eng., Part 2, pp. 959–975, 1980.

Pugh, D. T. and Woodworth, P. L.: Sea-Level Science: Understanding Tides, Surges, Tsunamis and Mean Sea-Level Changes, Cambridge University Press, 2014.

Rueda, A., Camus, P., Méndez, F. J., Tomás, A., and Luceño, A.: An extreme value model for maximum wave heights based on weather types, J. Geophys. Res.-Oceans, 121, 1262–1273, https://doi.org/10.1002/2015JC010952, 2016.

Rypkema, D. C., Horvitz, C. C., and Tuljapurkar, S.: How climate affects extreme events and hence ecological population models, Ecology, 100, e02 684, https://doi.org/10.1002/ecy.2684, 2019.

755 Scheffner, N. W., Borgman, L. E., and Mark, D. J.: Empirical simulation technique based storm surge frequency analyses, J. Waterway, Port, Coastal, Ocean Eng., 122, 93–101, https://doi.org/10.1061/(ASCE)0733-950X(1996)122:2(93), 1996.

Schellander, H., Lieb, A., and Hell, T.: Error structure of metastatistical and generalized extreme value distributions for modeling extreme rainfall in Austria, Earth and Space Science, 6, 1616–1632, https://doi.org/10.1029/2019EA000557, 2019.

Serinaldi, F. and Kilsby, C. G.: Rainfall extremes: Toward reconciliation after the battle of distributions, Water Resources Research, 50, 760 336–352, https://doi.org/10.1002/2013WR014211, 2014.

Smith, R. L.: Extreme value theory based on the r largest annual events, J. Hydrol., 86, 27–43, https://doi.org/10.1016/0022-1694(86)90004-1, 1986.

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 Resources Research, 53, 2833–2849, https://doi.org/10.1002/2016WR019426, 2017.

765 Songchitruksa, P. and Tarko, A. P.: The extreme value theory approach to safety estimation, Accident Anal. Prev., 38, 811–822, https://doi.org/10.1016/j.aap.2006.02.003, 2006.

Tawn, J. A.: An extreme-value theory model for dependent observations, J. Hydrol., 101, 227–250, https://doi.org/10.1016/0022-1694(88)90037-6, 1988.

Tawn, J. A. and Vassie, J. M.: Extreme sea levels: The joint probability method revisited and revised, in: Proc. Inst. Civ. Eng., Part 2, p. 770 429–442, 1989.

Tebaldi, C., Strauss, B. H., and Zervas, C. E.: Modelling sea level rise impacts on storm surges along US coasts, Environ. Res. Lett., 7, 014–032, https://doi.org/10.1088/1748-9326/7/1/014032, 2012.

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. S., 257, 107 418, https://doi.org/10.1016/j.ecss.2021.107418, 2021.

775 Volpi, E., Fiori, A., Grimaldi, S., Lombardo, F., and Koutsoyiannis, D.: Save hydrological observations! Return period estimation without data decimation, J. Hydrol., 571, 782–792, https://doi.org/10.1016/j.jhydrol.2019.02.017, 2019.

Von Mises, R.: La distribution de la plus grande de n valuers, 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. Dynam., 47, 3171–3190, https://doi.org/10.1007/s00382-016-3019-5, 2016.

780 Wahl, T., Haigh, I. D., Nicholls, R. J., Arns, A., Dangendorf, S., Hinkel, J., and Slangen, A. B.: Understanding extreme sea levels for broad-scale coastal impact and adaptation analysis, Nature communications, 8, 1–12, https://doi.org/10.1038/ncomms16075, 2017.

Woodworth, P. L. and Blackman, D. L.: Changes in high waters at Liverpool since 1768, Int. J. Climatol., 22, 697–7147, https://doi.org/https://doi.org/10.1002/joc.761, 2002.

Woodworth, P. L. and Blackman, D. L.: Evidence for systematic changes in extreme high waters since the mid-1970s, J. Climate, 17, 785 1190–1197, https://doi.org/10.1175/1520-0442(2004)017<1190:EFSCIE>2.0.CO;2, 2004.

Woodworth, P. L., Menéndez, M., and Roland Gehrels, W.: 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.

Zhang, K., Douglas, B. C., and Leatherman, S. P.: Twentieth-century storm activity along the U.S. east coast, J. Climate, 13, 1748–1761, https://doi.org/10.1175/1520-0442(2000)013<1748:TCSAAT>2.0.CO;2, 2000.

Zhang, W.-Z., Shi, F., Hong, H.-S., Shang, S.-P., and Kirby, J. T.: Tide–surge interaction intensified by the Taiwan Strait, J. Geophys. Res.-Oceans, 115, https://doi.org/10.1029/2009JC005762, 2010.

Zorzetto, E. and Marani, M.: Downscaling of rainfall extremes from satellite observations, Water Resour. Res., 55, 156–174, https://doi.org/10.1029/2018WR022950, 2019.

Zorzetto, E. and Marani, M.: Extreme value metastatistical analysis of remotely sensed rainfall in ungauged areas: Spatial downscaling and error modelling, Adv. Water Resour., 135, 103 483, https://doi.org/10.1016/j.advwatres.2019.103483, 2020.

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.