ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

D1.1.4
Review of spatial modelling of sea level extremes

Olle Räty and Marko Laine

19.12.2024

MAWECLI (MArine and WEather events in the changing CLImate as potential external hazards to nuclear safety) project in the SAFER2028 Programme

Finnish Meteorological Institute
P.O. Box 503
FI-00101 Helsinki, Finland
www.fmi.fi

Contents

Abstract

This report summarises a review of studies from the North Atlantic and Baltic Sea region, in which some form of spatial extreme value analysis was applied to storm surges. The 37 reviewed papers were categorised into 1) physics-based, 2) regional frequency analysis, 3) one-dimensional, 4) fully spatial and 5) satellite-based approaches. Studies based on physical modelling were the most numerous, likely because physical models directly provide spatial data for extreme value analysis without a need for using complex spatial statistical models. While there was a limited number of fully spatial modelling studies applied to storm surges in the study region, Bayesian hierarchical modelling with max-stable processes seemed the most promising for future research purposes. This approach allows for modelling the spatial dependence in the annual maxima and the extreme value distribution parameters at hand and also provides means to simulate spatial realisations of sea level extremes.

1 Introduction

In the Baltic sea region, reliable modelling of occurrence probabilities of sea level extremes is important for coastal planning, as the impacts of high sea level range from coastal erosion and habitat destruction to human losses (Rutgersson et al., 2021). The magnitude of extreme sea levels (ESLs) depends jointly on a multitude of physical phenomena such as moving low pressure systems, long term sea level changes and internal oscillations, which may or may not occur simultaneously. In the Baltic Sea, tides are relatively unimportant with some regional exceptions (Medvedev et al., 2016), but in many other parts of the world their inclusion is integral for proper analysis of sea level extremes.

Several strategies have been developed for estimating occurrence probabilities of high sea levels, both in the present and changing climate. Often the estimates are calculated using univariate approaches, in which only point-wise estimates of occurrence probabilities of, e.g., annual sea level maxima are provided. The most widely used approaches are the block-maxima approach, leading to the generalised extreme value distribution (GEV), and the peaks-over-threshold approach, using the generalised Pareto distribution (GPD) as a model for threshold exceedances. There are also other slightly different distributional choices based on the aforementioned ones such as the r-largest GEV distribution. An overview of these approaches is given in Coles (2001), but their main properties will also be summarised in this review.

A natural extension to univariate modelling of sea level extremes is to account for spatial (and temporal) variations in the extremes. The motivation of using spatial models is obvious. By taking the spatial dependence into account, estimates of occurrence probabilities and corresponding return levels are improved (less biased and less uncertain) over data-sparse regions. Most spatial modelling approaches can be used to calculate return level estimates of storm surges between tide gauge locations and can also directly simulate time series of sea level extremes in ungauged locations.

Extreme value analysis (EVA) studies can be roughly divided into two categories. The first one covers different physics-based modelling approaches, in which hydrodynamic simulations are performed to provide information about sea level variations. These are then used, potentially with additional covariate information, to analyse statistics of sea-level extremes (often using univariate extreme value analysis approaches) in the present-day and changing climate. The second category covers methods, in which EVA is performed on some form of sea level observations, although these approaches may additionally use model simulated quantities to facilitate spatial and temporal modelling of the extremes.

One approach to include spatial information to the extreme value modelling is to apply the so-called regional frequency analysis (RFA) (Hosking et al., 1997). In RFA, suitably standardised data is pooled over a region of interest according to certain rules in order to improve the reliability of the estimated probabilities for rare events. However, more attention has been recently given to more complex spatial models which explicitly model the spatial dependence of sea level extremes. A downside is that fully spatial models are often substantially more complex than the more traditional EVA approaches.

Statistical models for spatial extremes have been applied to observed and simulated data of many meteorological variables such as precipitation (e.g., Cooley et al., 2007; Sang et al., 2010; Reich et al., 2012). However, less work seem to be made with respect to sea level extremes. To advance work related to modelling of sea level extremes within the MAWECLI project framework, we make a systematic literature review of recent work on spatio-temporal modelling of sea level extremes in the Baltic Sea and North Atlantic coastal regions, the main focus being on the spatial modelling aspects. We first introduce the review framework, summarise the basic information of the reviewed articles and discuss the basics of EVA. We then give a more detailed overview of the reviewed approaches to performing EVA on sea level extremes and conclude with a summary of the main benefits and shortcomings of the different methodological approaches. We also briefly give suggestions to future research directions on spatial extreme value modelling within the MAWECLI framework.

2 The review framework

Articles were first collected and filtered based on a certain set of keywords from Google Scholar (`https://scholar.google.com/`), Clarivate Web of Science (`https://www.webofscience.com/wos/woscc/basic-search`) and Semantic Scholar (`https://www.semanticscholar.org/`). We only considered papers written in english and that had gone through a peer-review process. Some older potentially useful articles were unfortunately unavailable to us and thus, were not included here. We constrained the search to articles in which the analysis was geographically focused either in the Baltic Sea or the North Atlantic coastal region. However, we also included global-scale studies, if they covered the aforementioned regions. Furthermore, it was required that some form of statistical EVA be explicitly applied spatially or to spatial data in the paper. To further constrain the review, articles in which the wave height or tropical cyclones were exclusively analysed were omitted. Waves in particular have some data specific features that are out of scope of this review. However, this decision admittedly left out some articles, which would had likely been methodologically useful for our purposes. A final screening based on the references within the initial set of articles was made in an attempt to include missed articles to the extent possible.

A total of 37 articles matching the criteria of our review were found from the various literature sources. These articles were divided into the following five categories: 1) Physics-based approaches, 2) Regional frequency analysis, 3) one-dimensional spatial methods, 4) fully spatial statistical methods and 5) satellite-based approaches. There was a distinct lack of papers covering the Baltic Sea region in category 4, which lead to the decision to include a larger geographical region in the review. The reviewed articles and their main properties are listed in Table 1.

A basic summary of the approaches and methods applied in the reviewed articles is shown in Fig. 1. The most common approach among the reviewed articles (21 articles in total) was the application of EVA on physics-based simulations of extreme sea levels. The most popular distribution applied in these studies was the generalised Pareto distribution (nine articles) when inferring occurrence probabilities of

ESL events. In seven articles, either the GEV distribution or Gumbel distribution was used. The reason why the Gumbel distribution was favoured over the GEV distribution in so many articles, was likely due to its simple form, which facilitated parameter estimation from simulated data. Furthermore, in four out of 21 articles the r-largest GEV distribution was used to model the sea level extremes, and one also considered the exponential distribution.

The observation-based studies were further divided into five additional categories based on the spatial modelling approach taken in them. The regional frequency analysis (RFA) was the most popular approach and applied in five out of 16 articles. Bayesian hierarchical modelling (BHM) and one-dimensional regression-based models were both applied in three articles. Furthermore, there were two studies which used satellite altimeter data for spatial analysis and one in which stochastic modelling of ESLs was applied. Distributional choices for EVA were also more numerous than in the physical modelling studies. A total of eight different distributions were used, although not all of them were applied in spatial context. Overall, GEV (12 out of 15) and GPD (five out of 13) were the most popular choices, with a number of other distributions applied in one or two studies.

Table 1: A summary of the main features of the reviewed studies. The abbreviations are described in the main text.

	Author	Period	Region	Type	Method
1.	Rashid et al., 2024	1950–2017	CONUS	Statistical	GEV, Stochastic
2.	Bij De Vaate et al., 2024	1993–2021	Global	Statistical	GEV, Satellite
3.	Muis et al., 2023	1951–2050	Global	Physical (GTSM)	GPD, Exponential, GEV, Gumbel
4.	Li et al., 2023	1979–2018	Global	Physical (GTSM)	GPD
5.	Lorenz et al., 2023	1979–2018	Baltic Sea	Physical (GETM)	GEV, GPD
6.	Calafat et al., 2022	1960–2018	Western Europe	Statistical	BHM, GEV
7.	Andreevsky et al., 2020	1846–2011	Western Europe	Statistical	RFA, GPD
8.	Calafat et al., 2020	1960–2013	North Sea	Statistical	BHM, GEV
9.	Muis et al., 2020	1979–2017, 1976–2100	Global	Physical (GTSM)	GPD
10.	Beck et al., 2020	1966–2015	Nort-West Atlantic Coast	Statistical	BHM, GEV
11.	Muis et al., 2019	1988–2015	US North-Atlantic coastline	Physical (GTSM)	GPD
12.	Vousdoukas et al., 2018	2000–2100	Global	Physical (DFLOW FM, FES2014)	GEV, GPD

Continued on next page

Table 1: A summary of the main features of the reviewed studies. The abbreviations are described in the main text.

	Author	Period	Region	Type	Method
	Soomere et al., 2018	1961–2005	Baltic Sea	Statistical	GEV, Gumbel, 2-parameter Weibull
14.	Lobeto et al., 2018	1993–2015	US North-Atlantic coastline	Statistical	GEV, Satellite
15.	Frau et al., 2018	1846–2017	Western Europe	Statistical	RFA, GPD
16.	Wahl et al., 2017	1979–2014	Global	Physical (GTSR)	Gumbel
17.	Muis et al., 2017	1979–2014	Global	Physical (GTSR)	Gumbel
18.	Muis et al., 2016	1979–2014	Global	Physical (GTSR)	Gumbel
19.	Vousdoukas et al., 2016	1970–2005, 2010–2040, 2070–2100	Europe	Physical (Delft3D)	GPD
20.	Weiss et al., 2013	1915–2011	British Isles	Statistical	RFA, GEV, GPD
21.	Marcos et al., 2012	1950–1999, 1958–2001, 2000–2099	Bay of Biscay	Physical (HAMSOM)	GEV
22.	Gräwe et al., 2012	1960–2000, 2000–2100	Baltic Sea	Physical (GETM)	GEV, r-largest GEV, GPD
23.	Marcos et al., 2009	1950–1999 1958–2001, 2000–2099	Mediterranean, Iberian coast	Physical (HAMSOM)	r-largest GEV
24.	Bardet et al., 2011	1846–2008	North-East Atlantic coast	Statistical	RFA, GPD, Exponential, Mixed exponential
25.	Bernardara et al., 2011	early 19th century onward	Western Europe	Statistical	RFA, GPD
26.	Haigh et al., 2010	1900–2006	English Channel	Statistical	GEV, r-largest GEV, JPM, RJPM, SRJPM
27.	Marcos et al., 2009	1958–2001	Mediterranean, Iberian coast	Physical (HAMSOM)	GPD

Continued on next page

Table 1: A summary of the main features of the reviewed studies. The abbreviations are described in the main text.

Author	Period	Region	Type	Method
28. Wang et al., 2008	1961–1990, 1990–2002, 2031–2020	Irish waters	Physical (ROMS)	GEV
29. Butler et al., 2007a	1955–2000	North Sea	Physical (CSX)	r-largest GEV
30. Butler et al., 2007b	1955–2000	North Sea	Physical (CSX)	r-largest GEV
31. Bernier et al., 2006	1960–1999	North-West Atlantic	Physical (POM)	Gumbel
32. Meier, 2006	1961–1990, 1903–1998, 2071–2100	Baltic Sea	Physical (RCO)	Gumbel
33. Lowe et al., 2005	1961–2000, 2071–2100	British Isles	Physical (CSX)	GEV
34. Lowe et al., 2001	2006–2036, 2081–2100	British Isles	Physical (CSX)	Gumbel
35. Dixon et al., 1998	NA	British Isles	Statistical	1-d regression, GEV, RJPM
36. Dixon et al., 1992	1813–1988	British Isles	Statistical	1-d regression, GEV
37. Coles et al., 1990	1813–1988	British Isles	Statistical	1-d regression, GEV

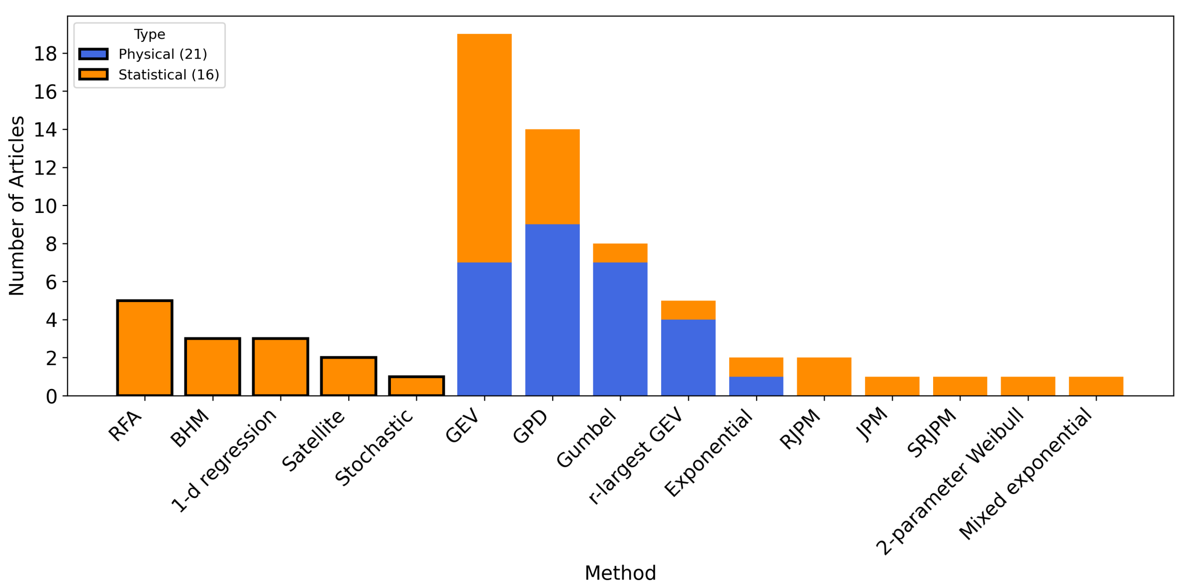

Figure 1: A summary of the methods used in the reviewed articles. The counts are shown separately for physical and statistical modelling articles. The first five boxes show the spatial approaches taken in the statistical modelling articles. The abbreviations are described in the main text.

3 Preliminaries

In this section, we briefly recap the basics of the extreme value theory following Coles (2001). Starting with the block maxima approach in the univariate case (i.e., no spatial dependence), let $Y_i = \{Y_1, \ldots, Y_n\}$ be a sequence of independent and identically distributed (i.i.d.) random variables. We are interested in the behavior of $M_n = \max\{Y_1, \ldots, Y_n\}$, for example the annual maximum sea level in case n is the number of observations over a year long period. The classical extreme value theory states that assuming there exist sequences of normalizing constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$\Pr\{(M_n - b_n)/a_n\} \to G(y) \text{ as } n \to \infty, \tag{1}$$

then the only suitable limiting distribution for G is the GEV distribution

$$G(y;\mu,\sigma,\xi) = \begin{array}{ll} \exp\left\{-\left[1+\xi\left(\frac{y-\mu}{\sigma}\right)\right]_+^{-\frac{1}{\xi}}\right\}, & \text{for } \xi \neq 0 \\ \exp\left\{-\exp\left[-\left(\frac{y-\mu}{\sigma}\right)\right]\right\}, & \text{for } \xi = 0. \end{array} \tag{2}$$

The three parameters $\mu \in \mathbb{R}$, $\sigma > 0$ and $\xi \in \mathbb{R}$ in Eq. (2) are the location, scale and shape parameter, respectively. The GEV tail behavior depends on the shape parameter ξ such that when $\xi < 0$ (Weibull), y has an upper limit at $\mu - \xi/\sigma$, whereas for $\xi = 0$ (Gumbel) and $\xi > 0$ (Fréchet), the tail is unbounded. Equation (2) can be inverted to obtain the GEV quantile function

$$y_p = \begin{array}{ll} \mu - \frac{\sigma}{\xi}\left[1 - \{-\log(1-p)\}^{-\xi}\right], & \text{for } \xi \neq 0 \\ \mu - \sigma\log\{1 - \log(1-p)\}, & \text{for } \xi = 0, \end{array} \tag{3}$$

In Eq. 3, y_p is the p-quantile corresponding to certain exceedance probability p, also known as the return level. The quantile function can be alternatively expressed using the definition of return period $T = 1/p$. For typical hydrological and meteorological observations, a block size of one year is often considered to be large enough for the GEV distribution assumptions to approximately hold, but other block sizes (e.g. monthly) have also been used.

To cope with the limited amount of data available for estimating the GEV distribution parameters, the so-called r-largest GEV was developed as an extension to the traditional GEV distribution. In this case, r largest values from each data block are retained instead of using only the largest one. The same assumptions hold, as in the case of the GEV distribution, and the same distribution parameters are estimated, using the joint density for the r largest values

$$h(y_1,\ldots,y_r;\mu,\sigma,\xi) = \exp\left\{-\left[1+\xi\left(\frac{y_r-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}\prod_{k=1}^{r}\frac{1}{\sigma}\left[1+\xi\left(\frac{y_k-\mu}{\sigma}\right)\right]^{-(1+1/\xi)}. \tag{4}$$

The benefit of the r-largest GEV is that more information about the extremes is included, which in principle facilitates the estimation of GEV distribution parameters. However, this is made with the expense of possibly deviating from the theoretical assumptions for too large values of r, which might lead to biases in the parameter estimates. More information about this approach is provided by Coles (2001).

Another commonly used approach is the so-called peaks-over-threshold approach, which is analogous to block-maxima approach. Denoting excesses over some large enough threshold u as $X_i = Y_i - u$ with $Y_i > u$, the limiting distribution for the threshold excesses is the generalised Pareto distribution (GPD)

$$H(x;\tilde{\sigma},\xi) = \begin{cases} 1-(1+\frac{\xi x}{\tilde{\sigma}})^{-1/\xi}, & \text{for } \xi \neq 0 \\ 1-\exp(-\frac{x}{\tilde{\sigma}}), & \text{for } \xi = 0, \end{cases} \tag{5}$$

where $\hat{\sigma} = \sigma + \xi(u-\mu)$ is the GPD scale parameter. The shape parameter ξ is equivalent to the GEV shape and defines the tail behaviour accordingly. The GPD quantile function corresponding to N-year return level is defined as

$$x_p = \begin{cases} u + \frac{\tilde{\sigma}}{\xi}[(Nn_y\lambda_u)^{\xi}-1], & \text{for } \xi \neq 0 \\ u + \tilde{\sigma}\log(Nn_y\lambda_u), & \text{for } \xi = 0 \end{cases}, \tag{6}$$

where n_y is the number of observations per year and λ_u is the probability of exceeding u. One of the main challenges in using GPD is how to select the threshold u, and there are several ad-hoc methods developed for selecting a suitable threshold value (e.g., Scarrott et al., 2012). Often, as low threshold as possible is chosen such that GPD is still a reasonable approximation to the data, so that the sample size is not too severely limited (Coles, 2001).

4 Spatial extreme value modelling approaches

We next discuss the various methods used in the reviewed extreme value modelling studies. The main weight is on describing the methodological choices rather than on the modelling results, but the most important ones are presented whenever they support the methodological discussion and help to understand the strengths and weaknesses of the method at hand.

4.1 Physics-based modelling

Physical modelling provides a compelling alternative to statistical spatial modelling approaches when analysing ESLs (Weisse et al., 2021). The main strength of physical models is that they, by definition, can simulate physically meaningfully spatio-temporal sea level processes and provide spatially coherent results; as Butler et al. (2007b) state, "from the statistical perspective the model output can effectively be regarded as spatially and temporally resolved storm surge elevation data". One important benefit of using physics-based models is that the contributions of different factors that drive ESLs can be evaluated in parallel with the statistics of ESLs. Furthermore, the limited availability of observed sea level extremes can be in principle alleviated by running model simulations over long time periods, although it is computationally costly to perform long-term simulations with complex physics-based models. Studies based on simulated sea levels typically, but not always, apply EVA independently on each grid box with the assumption that the model directly resolves spatial dependencies. Physical modelling is also required to provide covariate information for the extreme value models, when future changes in the sea level extremes are studied.

While there have been several physic-based modelling studies on extreme sea levels in the study region, many used the tail quantiles of the empirical distribution, when analysing the tail of behavior of ESLs. To keep the study focused, we restricted our review solely to those papers that explicitly included some form of theoretical extreme value analysis. This allowed to make some comparisons between the physics-based and statistical modelling approaches. Most of the physical modelling studies were made on a regional scale and provided location-specific information about extreme sea levels. Therefore, we review their main aspects on a regional basis.

The first physics-based extreme value analyses on the North Sea region and the British Isles were made by Lowe et al. (2001) and Lowe et al. (2005). They studied future changes in storm surges around the British Isles and the adjacent sea areas with the CSX model (e.g., Flather et al., 1998), using the Gumbel and GEV distribution for EVA. Model simulations in these early studies had a rather coarse resolution (35 km), which resulted with an underestimation of return level estimated with respect to the observations. Another climate study was conducted in the same region by Wang et al. (2008), who projected spatial changes of storm surge return levels on the Irish waters in the mid-21st century conditions by applying the standard GEV distribution to the regional oceanic model system (ROMS) simulations (Shchepetkin et al., 2005). The main benefit with respect to the earlier studies was the increased spatial resolution ($\sim$ 7 km) and added insights into the surge generating mechanisms. Note that in all these studies the five largest independent surges from each year were used at least in some parts of the study to fit the model instead of yearly block maxima, but it is not immediately apparent, whether the r-largest approach was used or not. All these studies underlined the importance of reliable forcing data.

The study by Butler et al. (2007b) was the first one, where a statistical spatial extreme value model was applied to simulated storm surges. They used a coarse ($\sim$35 km) resolution, 46 years long storm surge reconstruction generated with the CSX model over the North Sea region and fitted three variants of the r-largest GEV distribution to the data: one with grid-box specific parameters, one with a non-stationary model for the location and scale parameter and one in which these two parameters were allowed to vary both in space and time. They used the 20 largest independent annual values in their analysis, which is a rather large number compared to the other reviewed studies. The spatial and temporal dependence were modelled using a non-parameteric kernel approach that enforces local spatial and temporal smoothness in the two GEV parameters. Model uncertainty was assessed by bootstrapping (resampling with replacement) temporal samples from data, while keeping the spatial structure fixed. Butler et al. (2007b) showed that this model allowed to generate smooth spatial estimates of temporal trends in the location and scale parameter. However, a major shortcoming of this approach is that there is not an easy way to select an optimal kernel bandwidth for handling the spatial and temporal smoothness, when estimating the model parameters. Again, physical model deficiencies such as the coarse resolution were acknowledged to affect the results in certain areas. A similar study was performed by Butler et al. (2007a) who used the CSX model and the r-largest GEV in a similar manner to Butler et al. (2007b), with the main weight being on the decadal-scale storm surge variations.

In the Baltic Sea region, which is the region of most interest in the review, the first EVA study using physical modelling was made by Meier (2006). He used the Gumbel distribution, including uncertainty estimates based on bootstrapping, to infer the spatial pattern of 100-year return levels from an 11 km resolution hindcast and time-slice simulations run with the RCO model (e.g. Meier et al., 1999). The return level values were reasonably close to the observed values apart from the western Baltic Sea, Finnish archipelago and the end of the Gulf of Finland, where they were underestimated. The results showed that when changes in mean sea level were included, increases in the exceedance probability of extreme sea levels above 160 cm were seen over large areas on the westward-facing coastal regions of the Baltic Sea during the 21st century. Later, Gräwe et al. (2012) performed a more detailed, high-resolution (1 km spatial resolution) transient simulation with the General Estuarine Transport Model (GETM) (Burchard et al., 2002) over the western Baltic Sea. They compared three distributions, when modelling the storm surge height: GEV, r-largest GEV and GPD. By comparing return level estimates obtained from distribution fits to observations and model simulations at several tide gauge locations, they concluded that the r-largest GEV distribution with $r = 5$ provided a sufficient fit to the simulated data and also had the smallest uncertainty range. Relative biases in 10 to 50-year return level estimates were about $\pm 5\%$, which indicated that the high resolution model gave accurate results. Locally, the errors were still of the order of 10–20%, likely due to misrepresentation of local topography, although the limitations of atmospheric forcing could also explain part of the error. Over large areas, the GEV shape parameter was negative, which is in line with some observation-based statistical studies (e.g., Räty et al., 2023).

Soomere et al. (2018) studied various extreme value distributions on the southern coast of the Gulf of Finland and the Gulf of Riga, using RCO and RCA4 NEMO (Wang et al., 2015) simulations from 1961 to 2005. Both models were run at a 2 nautical mile (3.7 km) resolution, but with different temporal resolutions. These simulations were used to evaluate spatial variations in the GEV, Gumbel and the 2-parameter Weibull distribution parameters. Both the location (of the GEV and Gumbel distribution) and scale parameter increased towards the end of the Gulf of Finland, whereas the shape parameter (of the GEV and Weibull distribution) stayed relatively constant throughout the studied coastal region. Soomere et al. (2018) also compared results from various parameter estimation procedures, when estimating the distribution parameters. The results showed that the parameter values depended to certain extent on the chosen estimation procedure, but the used physical model also affected the parameter estimates.

As an extension to the work by Gräwe et al. (2012), Lorenz et al. (2023) analyzed the capability of GETM to simulate ESLs in the whole Baltic Sea. The model was run for 40 years with various forcing data sets to evaluate uncertainties related to the atmospheric forcing. Both the GEV distribution and GPD were fitted to the model simulations and the 30-year return level was calculated from the fitted distributions. The results showed that the return level estimates were relatively independent of the EVA method used, with a well-known spatial pattern of higher return levels on the eastern side of the Baltic sea and on the west facing coastal regions elsewhere. However, there were large spread in the return level estimates between the forcing data sets, underlining the fact that uncertainties in the return level estimates are strongly linked with the representation of atmospheric forcing.

Over the southern Europe, Marcos et al. (2009) and Marcos et al. (2011) Performed storm surge simulation studies around the Iberian coastal region and the Mediterranean basin with the HAmburg Shelf Circulation Model (HAMSOM) model run on $1/6° \times 1/4°$ resolution (Ratsimandresy et al., 2008). Marcos et al. (2009) evaluated the model's capability to reproduce observed sea level extremes in the present-day conditions, using GPD fitted to five largest values per year. As in the previous studies, the model-estimated return levels were slightly underestimated in many locations compared to observations. Marcos et al. (2011) calculated spatial projections of storm surge return levels under the 21st century climatic conditions, using seven non-stationary (location and scale parameter) variants of the r-largest GEV distribution with $r = 5$, or using only the annual maxima in regions, where the r-largest GEV fit was poor. They also performed EVA similarly for the negative storm surges. There were marked spatial variations in the parameters of the best performing model (the location parameter modelled as a linear function of time and the North Atlantic Oscillation index (NAO)), which were apparent also in the spatial return level estimates. Marcos et al. (2012) performed an interesting case study, in which the same modelling setup with Marcos et al. (2011) was used, but with a quantile mapping post-processing step to correct for model underestimation, when projecting changes in various return levels under two future emission scenarios around the end of the Bay Of Biskay. In addition to a standard EVA, they used the simulated changes in the 50-year return level to estimate the spatial extent of coastal flooding in the city of Bilbao.

The first European scale analysis of storm surges was performed by Vousdoukas et al. (2016), who fitted the GPD to the Delft3D storm surge model simulations (Deltares, 2014) run with several atmospheric and climate model forcings. Model validation against tide gauge observations along the European coast indicated a relatively good overall performance, with some location-specific over- or underestimation of low and high extremes. By applying EVA to an ensemble of storm surge projections, they were able to derive changes in the simulated return levels up to 500 years, which showed substantial increases in the North Sea and Baltic Sea region.

On the western North-Atlantic coast, the first physics modelling-based extreme value analysis of storm surge was conducted by Bernier et al. (2006). They calculated the 40-year spatial return levels by fitting the Gumbel distribution to a 40-year long hindcast made with the Princenton Ocean Model (POM) simulation (Mellor, 1998) that was run with an approximately 9 km resolution. Validation against a set of tide gauge fits showed that the modelled 40-year return level estimates were reasonable with a slight underestimation in some locations. They attributed the underestimation to the limited resolution of forcing (wind and pressure) fields, which was insufficient to resolve stronger tropical systems in the southern parts of the model domain. Also the model resolution and model limitations in simulating local processes affected the results. Muis et al. (2019) studied the spatio-temporal patterns of ESLs and their driving factors over the western parts of the North Atlantic region. They used a 28-year long sea level reanalysis (a combination of past observations with model simulations) generated with the Global Tide Surge Model (GTSM) (Muis et al., 2016) and fitted the GPD to storm surges above the 99th percentile threshold. Due to the short time period used, they focused on return levels only up 10 years. Analysis of the driving mechanisms and their contributions to ESLs showed that extra-tropical cyclones had the largest contribution in the northern parts of the study domain, with tropical cyclones having a more local, but still significant contribution particularly to the most extreme sea levels in the southern parts of the domain.

On a global scale, the first physics-based reanalysis of extreme sea levels (GTSM) was provided by Muis et al. (2016). They fitted the Gumbel distribution to the simulated storm surge from 1979 to 2014 and concluded that, in regions where extra-tropical cyclones are the main driving factor of storm surge, the modelled return level estimates were reasonable, although there was a general tendency to underestimate them in comparison to the observations. The same modelling framework was later adapted in other studies (Muis et al., 2017; Wahl et al., 2017), which confirmed the underestimation of return levels with respect to the observations. Muis et al. (2020) produced a newer global ocean reanalysis in which the biases in the 10-year return levels, inferred using the Gumbel distribution, were slightly smaller compared to Muis et al. (2016). Yet in another study, Muis et al. (2023) performed a global extreme value analysis of storm surges using the GTSM model forced by five climate models from the latest Coupled Model Intercomparison Project (CMIP6) with the GPD fitted above the 99th percentile threshold (other distributions were initially evaluated). Climate model simulations had positive biases in the estimated return levels in comparison to a baseline simulation at high latitudes and particularly in semi-enclosed basins. This bias was attributed to the too intensive storms simulated by the climate models in these regions.

Vousdoukas et al. (2018) provided global probabilistic projections of ESLs, including contributions from mean sea level, tides, and the combined effect of storm surges and wind-waves in their analysis. For storm surges, a similar modelling setup as in Muis et al. (2016) was used. Their EVA was based on a modification to the GEV distribution and GPD under non-stationary conditions by Mentaschi et al. (2016). They applied the following time-varying normalisation to the observations

$$x(t) = \frac{y(t) - T(t)}{S(t)}, \tag{7}$$

where $T(t)$ and $S(t)$ are the trend and slowly varying standard deviation of the original time series

$y(t)$. However, it was not apparent, whether Vousdoukas et al. (2018) applied either the block-maxima or peaks-over-threshold approach in their study.

A sophisticated approach to generate global, spatially dependent synthetic extreme sea levels was demonstrated by Li et al. (2023). They used 40 years of global sea level reanalysis data produced with GTSM to generate a set of synthetic sea level extremes and spatially dependent return levels along the global ocean coasts. Their main aim was to generate a large enough number of events that allowed a robust estimation of rare sea level events and their uncertainty. Analysis locations were clustered up to 30 small clusters within ten disjoint coastal regions based on POT time series, using the 95th percentile as a threshold. Li et al. (2023) then fitted a multivariate conditional exceedance model of Heffernan et al. (2004) to the reanalysis to model the joint distribution of 3-day water level maxima. The modelling approach consisted of first defining the marginal distributions on each analysis location and then calculating pair-wise dependence between the locations within each cluster. More formally, let $F_i(X_i)$ with $i = 1, \ldots, d$ be the marginal distribution of 3-day maximum water levels X_i on analysis location i among d locations within a target cluster. Each marginal distribution was estimated as a mixture distribution by fitting the GPD above a location-specific percentile threshold u_i and an empirical distribution below u_i. Then, the individual distributions $F_i(X_i)$ were transformed to a common scale using the Laplace distribution (Keef et al., 2013b). This procedure transforms both the upper and lower tail of the distribution to exponential. Next, pair-wise dependencies were estimated using

$$\mathbf{Y}_{-i}|Y_i = \mathbf{a}Y_i + Y_i^{\mathbf{b}}\mathbf{Z}_{-i}, \ \mathrm{y} > \nu, \tag{8}$$

where $\mathbf{Y}_{-i}$ is a vector of all marginal distributions apart from Y_i (the target location) and ν is a threshold above which the pair-wise dependence is calculated. The parameter vectors $-1 < \mathbf{a} < 1$ and $\mathbf{b} < 1$ describe the strength and variation of the dependence, respectively. Vector $\mathbf{Z}_{-i}$ contains $d-1$ residuals independent of Y_i. A quantile-based estimation procedure for $\mathbf{Z}_{-i}$ is used as described by Keef et al. (2013a). The calculation of pair-wise dependence was repeated for each pair of locations within each cluster.

In the final step, synthetic, spatially dependent sea level events were generated stochastically using the multivariate conditional exceedance model. For doing this, Li et al. (2023) estimated the distribution of annual event counts within each cluster from the 99th percentile exceedances of the 3-day maximum sea level time series separately at each analysis location with a kernel-based approach. By sampling from the kernel distribution, time series of event counts corresponding to 10000 years of extreme sea level events were obtained. The sampled events were distributed within each cluster such that the proportion of extreme sea level events at each location matched the empirical estimates calculated from the scaled distribution $\mathbf{Y}$. The conditional model was sampled with the correct proportions and the obtained synthetic ESLs were finally used to calculate return levels empirically from the sample. Li et al. (2023) argued that the empirical estimation approach does not suffer from the same limitations as the theoretical models (e.g. difficulties in estimating the shape parameter) and that the sample size is sufficient to estimate 1000-year return levels. It should be noted, however, that their model inherently assumes that the tail of the source data distribution follows the GPD. Simulation experiments showed that their model provided good performance in comparison to the reanalysis data set, with slight underestimation of return level values for long return periods. Furthermore, a comparison against historical events showed that the model is able to simulate realistic spatial dependence.

To conclude, physics-based modelling allows to gain physically consistent spatial information about ESLs and gives insights in the physical factors governing them. Furthermore, future changes in the storm surge extremes can be analysed. However, in many cases the limited temporal span of simulations restricted the analysis of return levels such that relatively commonly occurring sea levels (e.g. 10-year return level) were targeted. Furthermore, studies often incorporated relatively simple extreme value analysis methods such as the GEV distribution and GPD. The most comprehensive EVA models reviewed were the ones applied by Butler et al. (2007b), who used a non-parametric model to capture spatio-temporal variations in the location and scale parameter of the GEV distribution and Li et al. (2023), who implemented a stochastic sampling scheme, which was capable to generate very long, realistic, synthetic time series of ESL events.

4.2 Regional frequency analysis (RFA)

A form of spatial analysis in which all observations are pooled together over a region is called regional frequency analysis (RFA), as described in its original form by Dalrymple (1960). For a formal description, let us have N tide gauges with n_i observations available at each gauge i so that the overall observations are Y_{ij}, for $j = 1, \ldots, n_i$. Also, let us define the local quantile function corresponding to a cumulative probability F as $Y_i(F)$, $0 < F < 1$. The regional frequency distribution $\hat{Y}(F)$ is defined as $\hat{Y}(F) = Y_{ij}(F)/\mu_i$, where μ_i is a site-specific index-flood that describes the site specific traits of the observed frequency distribution. Not that μ_i should not be confused with the location parameter of the GEV distribution. Alternatively μ_i is also know as the surge index (Bernardara et al., 2011) or the local index. We will use latter term hereafter (Weiss et al., 2013).

The main assumption in RFA is that the frequency distributions belonging to a certain group of sites are similar apart from the scaling factors μ_i. There are several ways to define μ_i. The distribution mean or some other central statistic is often used, although distribution quantiles or some high sea level value (Bardet et al., 2011) have also been tested for this purpose. Other physical quantities can also be considered when calculating the local index, which in principle allows to make some estimations at un-gauged site.

In practice, the calculation of regional frequency curves in RFA proceeds according to Hosking et al. (1997) as follows:

1. Data is first screened for homogeneity and inconsistencies are eliminated
2. Observations within the region of interest are categorised into approximately homogeneous groups according to their distributional similarity.
3. The observations are then standardised using the local index.
4. A suitable probability distribution is chosen based on goodness-of-fit tests and then fitted to observations either at each site or to the whole regional data after which the results are pooled together to get the regional frequency distribution.

A useful concept related to RFA is the effective duration D_{eff}, which describes the equivalent record length obtained by pooling data together from multiple sites. The value of D_{eff} strongly depends on inter-site dependence, i.e., the number of independent observations and is somewhere between the length of the longest tide gauge record (perfect inter-site dependence) and the total length of the pooled time series (no inter-site dependence). Weiss et al. (2014b) provide a formulation over N tide gauges as $D_{eff} = \psi \sum_{i=1}^{N} d_i/N$, where d_i is the length of time series at tide gauge i and $\psi = \lambda_r/\lambda$, $\psi \in [1, N]$ is a measure of inter-site dependence given as the ratio of average annual number of storms within a region λ_r to the average annual storminess at individual sites λ. Thus, D_{eff} directly summarises the benefit of pooling the observations with RFA.

A critical part of RFA is the assessment whether the observations from a certain region are homogeneous enough. Various ways for this have been suggested in the literature. Bernardara et al. (2011) tested regional homogeneity using a regional heterogeneity measure H and a local discordancy measure D_i. Both measures are based on the so-called L-moments, which provide an alternative system to describing the distribution moments. The basic idea of the heterogeneity measure H is to compare the dispersion of certain sample L-moment ratios of individual tide gauges with the sampling variability of a homogeneous group of tide gauges. A homogeneous region is simulated from a Kappa distribution, whose parameters are estimated using the regionally weighted L-moment ratios. The fitted distribution is then used to generate a large sample of synthetic observations from which the sampling variability is estimated. Hosking et al. (1997) provides guidelines how large a dispersion is allowed in the local L-moment ratios for a region to be considered homogeneous. On the other hand, Discordancy measure D_i measures whether any single tide gauge is too much in discordance with the other tide gauges within a region, using again the L-moments. As such, it can also be used for screening gross outliers before a homogeneous region is defined from the study sites. More information about both measures can be found from Hosking et al. (1997).

One additional difficulty in using RFA with storm surges is the possible temporal dependence between the gauges, which reduces the effective sample size of the pooled data. This aspect is often taken into account rather crudely due to the difficulty to distinguish between physically independent events (i.e., are there one or more storms causing high sea levels within a region over a short time window) and finding methods to cope with the inter-site dependencies. The simplest approach is to leave out other observations other than the highest one within a certain time window (Bernardara et al., 2011; Bardet et al., 2011), but doing this useful data will likely be lost. Weiss et al. (2014b) provide a theoretical framework for accounting for inter-site dependencies in the POT framework for extreme wave heights.

Some studies have also tried to automatize the definition of homogeneous regions. The framework proposed by Weiss et al. (2014a) includes a clustering method based on the concept of a storm footprint, which can be used to assess the strength of inter-site dependencies. Another approach was recently suggested by Andreevsky et al. (2020) for defining statistically and also physically homogeneous regions by using the extremal coefficient between two sites as $\rho(X, Y) = \lim_{n\to\infty} P[X > q_1 | Y > q_2]$, where q_1 and q_2 are some extreme surge quantiles of site specific random variables X and Y. The extremal index stems from multivariate extreme value theory as a measure for the dependence between the extremes.

As there are various definitions for the local index, there is a certain level of subjectivity in selecting the local index. Some guidelines are given by Hosking et al. (1997). Weiss et al. (2013) performed a comparison of four different methods for calculating the local index. They showed that the distribution mean tends to provide reasonably robust results for local scaling, whereas in case of asymmetric distributions, other indices could also be considered.

We now summarise the main findings from the reviewed articles. The earliest found study which performed RFA on storm surges was made by Bernardara et al. (2011) along the French Atlantic coast on 18 tide gauges, using the GPD. They concluded that although GPD did not provide a completely satisfactory fit to the regional data, the fit was much better compared to the site-specific fits due its capability to provide higher return levels. Also the uncertainty in the return level estimates was reduced.

Bardet et al. (2011) performed RFA also using long-term tide gauge time series from the French Atlantic coast. They compared a mixed exponential distribution with an exponential distribution and the

GPD in providing regional return level estimates, and concluded that the mixed exponential distribution provided the best fit to the data. In line with the study of Bernardara et al. (2011), their results showed that the regional 1000-year return level estimates were consistently larger at tide gauge locations than those obtained with the site-specific fits.

Weiss et al. (2013) provided guidelines for selecting a suitable local index under various levels of inter-site dependence, regional heterogeneity and distributional asymmetry. Out of the four tested local indices, they suggested of using the distribution mean as the local index, when the region is sufficiently homogeneous and the local distributions are symmetric. In the presence of asymmetry or for slightly heterogeneous observations, other indices such as the location parameter of the fitted (GEV or GPD) distribution could also be considered. They also demonstrated the differences on return level estimates between the tested local indices in a case study, which included 16 tide gauges from the British coastal region. The results showed that choosing a sub-optimal local index has a deteriorating effect on the estimated return levels. Overall, there is no a unique way to define the local index and improper selection can have an adverse effect on the return level calculations (Weiss et al., 2013).

Frau et al. (2018) attempted to include historical storm surges outside the continuous observations to the RFA, using the GPD as the distributional model. Their approach was based on the assumption that the average occurrence frequency of storms λ, which corresponds to a certain high quantile threshold, has stayed constant over time. They then defined the concept of credible duration D_{cr}, which is in essence D_{eff} that takes into account the historical duration due to the historical observations. For example, a single historical observation included with the threshold $\lambda = 0.1$ would correspond to an increase of ten years in D_{cr}. It should be noted that the notion of D_{cr} is only useful for the peaks-over-threshold approach. By assuming a constant λ over the whole observational period, they were able to include historical records in their RFA approach. Using a large number of tide gauges from the French and British coastal region, they demonstrated that the inclusion of non-continuous historical observations lead to an increase in the credible duration and higher return level estimates in comparison to the case, where the historical observations were not used.

Lastly, Andreevsky et al. (2020) described a method for defining homogeneous regions in a more physically meaningful matter, using the concept of empirical spatial extremogram (ESE). They used ESEs to construct homogeneous regions for three target sites in France and showed that the defined regions were more homogeneous than those provided by Weiss et al. (2014a), although the 1000-year return level values were relatively similar in both studies.

While the aforementioned studies have illustrated that pooling observations with RFA increases the robustness of the estimated occurrence probabilities of extreme sea levels, when the gauge sites are reasonably homogeneous, it has some shortcomings. Most importantly, RFA does not account for dependencies between separate regions and does not easily allow using covariate information in the extreme value modelling step, although this has been suggested in the literature (Hosking et al., 1997). Also, choices regarding the various steps and parameters in RFA pose some challenges for the practical use of this method in real-world cases.

4.3 One-dimensional spatial modelling

A natural extension to the univariate, point-wise modelling is to handle the coastal region of interest in a one-dimensional manner, for example using the distance and other covariates along the coast of the target region as predictors. Compared to full-fledged spatial modelling, the one-dimensional modelling approach is substantially simpler, as there is no need to design and model complex two-dimensional fields of extremes and their dependence structures. Various one-dimensional approaches have been suggested in the literature. Here, we discuss their main points in a chronological order.

One of the first efforts to model sea level extremes with a one-dimensional spatial model was made by Coles et al. (1990), who modelled the GEV distribution parameters using a parametric model, with the distance along the coast of Britain as a covariate. They also tried to capture the underlying surge and tide-surge interactions in their model formulations. A multivariate extreme value distribution was constructed assuming that i) sites are conditionally independent such that for the neighbouring sites i, j and k, the annual maxima Y_i and Y_k are independent given Y_j and ii) the dependence between the neighbouring sites can be modelled using a logistic dependence structure. These assumptions lead to the following bi-variate probability density between two tide gauges (Coles et al., 1990)

$$f(y_1, y_2) = (\sigma_1\sigma_2)^{-1}\tilde{y}_1^{(r-k_1)}\tilde{y}_2^{(r-k_2)}(\tilde{y}_1^r + \tilde{y}_2^r)^{-2+1/r}\left\{(\tilde{y}_1^r + \tilde{y}_2^r)^{1/r} + r - 1\right\}\exp\left(-(\tilde{y}_1^r + \tilde{y}_2^r)^{1/r}\right) \quad (9)$$

with $\tilde{y}_i = [1 - k_i(y_i - \mu_i)/\sigma_i]^{1/k_i}$, for $i = 1, 2$. Note that Coles et al. (1990) used the definition $k = -\xi$. The strength of the dependence is governed by $r \geq 1$, such that when $r = 1$ the two sites are completely independent and when $r \to \infty$ they are completely dependent. The GEV parameters were modelled at N sites as a function year j and distance d_i with respect to a reference position as

$$\mu_i = \alpha_i + j\beta_i + \sigma_i \ln \tilde{I}(\sigma_i) \quad (10)$$

$$\sigma_i = \exp(\beta_0 + \beta_1 d_i + \beta_2 d_i^2 + \beta_3 d_i^*) \quad (11)$$

$$k_i = \gamma_0 + \gamma_1 d_i + \gamma_2 d_i^* \quad (12)$$

for $i = 1, \ldots, N$. In Eq. 10, α_i and β_i are the regression coefficients to be estimated and $\sigma_i \ln \tilde{I}(\sigma_i)$ describes both the tidal effect and the relationship with σ_i. The details of $\tilde{I}(\sigma_i)$ can be found from Coles et al. (1990). This shows that the location parameter does not directly depend on the distance along the coast. The scale (Eq. 11) and shape (Eq. 12) parameters have a second and first order polynomial dependence on d_i, respectively. Furthermore, the covariate d_i^* is the distance from an estuary mouth, used if the tide gauge is located within an estuary. Model diagnostics inspected over eight coastal stretches showed that in reality, a simplified version of the model usually provided a better fit to the data than the complete model. In some locations, the distance from the estuarine mouth was found to explain variations in the scale and shape parameter values.

Assuming that the marginal distribution G_i at tide gauge i is a GEV distribution such that $Y_i \sim \text{GEV}(\mu_i, \sigma_i, \xi_i)$ for $i = 1, \ldots, N$, Coles et al. (1990) tested the following spatial multivariate extreme value distribution

$$\Pr\{Y_i \leq y_i : i = 1, \ldots, N\} = \left[\prod_{i=1}^{N} G_i(y_i)\right]^{B(\mathbf{w})}, \quad (13)$$

where $B(\mathbf{w})$ is a dependence function and $\mathbf{w} = \{w_1, \ldots, w_N\}$ a vector of weights, whose components are defined as $w_i = \ln G_i(y_i)/\ln[\prod_{j=1}^{N} G_j(y_j)]$s for $i = 1, \ldots, N$. A logistic model described in more detail in Coles et al. (1990) was used for $B(\mathbf{w})$ with dependence the on semi-diurnal tidal phase and distance between the tide gauges. The main motivation for using this model was to test how strongly spatial dependence of ESLs is related to the propagation of storm surge along the coastline while simultaneously accounting for the tidal phase. Coles et al. (1990) concluded that at short distances surge progression and topography might be the dominant factors explaining the inter-site dependence.

Dixon et al. (1992) extended the model of Coles et al. (1990) to account for spatial dependence in the temporal trend of the location parameter of the GEV distribution. They used a local-linear weighted least-squares estimator to estimate the trend parameter β_i from Eq. 10 on site i as $\beta_j = \gamma_{i,0} + \gamma_{i,1} d_j + \epsilon_j$, for all $j \in N_m(i)$, where the neighbourhood of site i is defined as $N_m(i) = j : i - m < j < i + m$ and ϵ_j are zero-mean random variables with variance ν_j^2. The distance d_j was defined circularly over the whole coastal region. The local trend estimates were then smoothed using a Gaussian kernel function.

Later, a similar strategy was adopted by Dixon et al. (1998), who used a kernel regression approach to derive spatial estimates of distribution parameters between the tide gauge locations in the British coastal region. Their method was built upon the so-called joint probability method (JPM) (Pugh et al., 1980) and its revised version (RJPM) (e.g., Tawn et al., 1989), denoted as the spatial RJPM (SRJPM) hereafter. The model for the probability of sea level Z_t at time t to exceed some high value z was

$$H(z) = \Pr\left(\max_{1 \le t \le T} Z_t \le z\right) \approx \left\{\prod_{t=1}^{T} F_{X|Y}(z - X_t)\right\}^{\theta_Z}, \tag{14}$$

where $0 < \theta_Z \le 1$ is the extremal index that measures the degree of clustering of the extremes and $F_{X|Y}$ is the conditional distribution function of the hourly surge Y with respect to the tidal level X. This model aimed at accounting for tide-surge interactions, which might be important in shallow water regions. The model can also be applied in case tide and surge are independent.

Starting from the tide-independent case, Dixon et al. (1998) assumed that the surge time series followed the GEV distribution $\mathrm{GEV}(\mu_Y, \sigma_Y, \xi_Y)$, but used a non-homogeneous Poisson process to estimate the parameters, as this allowed to use all hourly (declustered) surge values in the parameter estimation. For the tide-surge interaction case, Dixon et al. (1998) first normalised the hourly surge time series using $S_t = (Y_t - a(X_t))/b(X_t)$. This stabilises the surge time series (reduced non-stationarity) and allows to use the same parameter estimation procedures as in the tide-independent case by replacing the original GEV parameters with

$$\mu_Y(X_t) = \mu_S b(X_t) + a(X_t),\ \sigma_Y(X_t) = \sigma_S b(X_t) \text{ and } \xi_Y(X_t) = \xi_S, \tag{15}$$

where μ_S, σ_S and ξ_S denote the parameter estimates obtained from the normalised surge time series and $a(X_t)$ and $b(X_t)$ are parameters specified from data. Both the tidal component and the surge conditional on tides were estimated spatially in order to calculate return level estimates for the joint distribution. The parameters related to the surge distribution at any coastal distance d were calculated from the tide gauge specific parameter estimates using a univariate weighted kernel regression approach (Dixon et al., 1998). The results showed that the spatial model extension provided less biased estimates compared to a naive alternative in which the return levels were directly interpolated between the data sites. The spatial model by Dixon et al. (1998) was further compared by Haigh et al. (2010) against four indirect methods (GEV, r-largest GEV, JPM and RJPM) on the British coast. Their main conclusion was that SRJPM overestimated return levels in locations, where there were only short sea level time series available, but this mismatch could have been caused by the differing length of data used to estimate the SRJPM parameters.

More recently, Räty et al. (2023) used a flexible, non-parametric models to capture the distance dependence of the GEV parameters along the Finnish coast. They used a hierachical Bayesian modelling approach with both penalised splines and Gaussian processes in a regression setting, when estimating the location and shape parameters of the GEV distribution along the coast. Their results showed that the return level estimates were less uncertain when including distance dependence in their model compared to single-site fits.

To summarise, while the one-dimensional extreme value modelling approaches provide an improvement over univariate approaches that assume independence between the tide gauges, their use was earlier motivated at least partly by the fact that the computational tools and modelling capabilities were not matured enough to enable full spatial modelling. Moreover, many of these methods estimate some of the model parameters using ad hoc methods without clear guidelines on how to choose the values for these parameters.

4.4 Full spatio-temporal modelling

Models that fully capture the spatio-temporal nature of ESLs have recently been started to apply to both observed and modelled sea level extremes. Most of the reviewed methods were based on a Bayesian hierarchical modelling (BHM) approach with different techniques to describe the spatial process. One study stochastically generated synthetic sea level extremes from a spatial process for further analysis. The small number of found articles highlights the fact that this type of models have rarely been applied to sea level extremes in the study region.

The earliest found study was conducted by Beck et al. (2020), who used a copula-based approach to spatially model storm surges on the Atlantic coast of Canada, using incomplete observations from 21 buoys over the years 1966–2015. If $\mathbf{Y} = \{Y_1, \ldots, Y_N\}$ with $N = 21$ is a random vector of observed annual maximum sea level and assuming that the observations followed the GEV distributions, the posterior distribution of all model parameters $\Delta = (\Theta, \eta, \xi)$ was described by Beck et al. (2020) using the Bayes' rule as

$$p(\Delta|\mathbf{Y}) \propto p(\mathbf{Y}|\Theta, \xi)p(\Theta|\eta, \xi)p(\eta, \xi). \tag{16}$$

In Eq. 16, $\Theta = (\theta_1, \ldots, \theta_N)$ with $\theta_i = (\mu_i, \sigma_i)$ for $i = 1, \ldots, N$ contains the GEV location and scale parameters of marginal distributions at locations $\mathbf{s}_i$ and η is a vector of model hyper-parameters. Thus, the hierarchical model is described by the three terms on the right-hand side of Eq. 16, which denote the data layer, process layer and the prior layer, respectively.

In Beck et al. (2020), the joint distribution H of observations $\mathbf{Y}$, given parameters Θ and ξ (data layer), was modelled following Sklar's theorem. Given a copula C and marginal distributions G_i for $i = 1, \ldots, 21$ that were assumed to be GEV distributions, the joint distribution was expressed as

$$H(y_1, \ldots, y_{21}) = C\left[G_1(y_1), \ldots, G_{21}(y_{21})\right] \tag{17}$$

There are various parametric forms available for the copula C. Beck et al. (2020) used Student's t copula as their model for C, with its correlation matrix being estimated using the exponential correlogram. The authors noted that it is often difficult to work with copulas in high dimensional setting, as is the case in their article. Student's t copula was chosen, as it allows to estimate the joint distribution even if there are data missing from some of the buoys and as it is capable to capture tail dependence even though it is not an extreme value copula.

In the process layer, the location and scale parameter of individual GEV distributions were modelled using latent Gaussian processes as

$$\boldsymbol{\mu} \sim \mathcal{N}_{21}(\mathbf{X}_{\boldsymbol{\mu}}\beta_{\boldsymbol{\mu}}, \tau^2_{\boldsymbol{\mu}}\Sigma_{\boldsymbol{\mu}}) \quad \text{and} \tag{18}$$

$$\boldsymbol{\Phi} \sim \mathcal{N}_{21}(\mathbf{X}_{\boldsymbol{\Phi}}\beta_{\boldsymbol{\Phi}}, \tau^2_{\boldsymbol{\Phi}}\Sigma_{\boldsymbol{\Phi}}). \tag{19}$$

In this equation, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{21})$ and $\boldsymbol{\Phi} = (\ln(\sigma_1), \ldots, \ln(\sigma_{21}))$ and the design matrices $\mathbf{X}_{\boldsymbol{\mu}}$ and $\mathbf{X}_{\boldsymbol{\Phi}}$ contain p covariates (and the intercept) for the site-specific GEV location and shape parameters, with vectors β_{μ} and β_{Φ} containing the associated coefficients. Furthermore, Σ_{μ} and Σ_{Φ} are the correlation matrices with variance parameters τ_{μ}^2 and τ_{Φ}^2 that are assumed to be constant over the whole domain. Sea-level pressure was included as the only physical covariate in the model. In the parameter layer, prior distributions were specified for ten hyper-parameters. To simplify the model, Beck et al. (2020) assumed that the priors were mutually independent and non-informative (i.e. wide) with suitable limits added to them to avoid non-identifiability issues, which are common in Gaussian process models. A custom Markov Chain Monte Carlo algorithm was used to sample from the posterior parameter distribution.

Comparison of return level estimates obtained with single-site GEV fits and the BHM showed that the values were systematically higher for the BHM, although this could be attributed to the choice of using the median as the summary statistic, which deviates from the mean for non-symmetric distributions. However, uncertainty ranges were markedly larger for the single-site fits and covered unrealistic, negative values in some locations, whereas the BHM was able to provide more meaningful uncertainty ranges. The latent process description of the GEV parameters readily allowed to interpolate the parameters to ungauged locations, and the surge process could be simulated using the copula model to generate synthetic storm surge realisations. This is a major advantage over the one-dimensional models discussed in the previous section.

An alternative approach was described by Calafat et al. (2020), who applied a hierarchical max-stable model to generate a probabilistic reanalysis of ESLs in the North Sea region. A max-stable process is an infinite-dimensional generalisation of the univariate GEV distribution. Following Davison et al. (2015) and Schlather (2002), let $\{Y_i(\mathbf{s})\}$ for $i = 1, 2, \ldots, n$ be i.i.d. replicates of a spatial random fields on $\mathbf{s} \in \mathbb{R}^d$. A spatial random process $Z(\mathbf{s})$ is a max-stable process, if there exist continuous functions $a_n(\mathbf{s}) > 0$ and $b_n(\mathbf{s})$ such that

$$Z(\mathbf{s}) = \lim_{n\to+\infty} \frac{\max_{i=1}^{n} Y_i(\mathbf{s}) - b_n(\mathbf{s})}{a_n(\mathbf{s})}. \tag{20}$$

By this definition, the univariate marginal distributions of $Z(\mathbf{s})$ are also GEV distributions (Eq. 1). Many of the parametric forms for the max-stable process have been derived from the spectral representation of Haan (1984), which facilitates the use of max-stable processes in real-world problems. Some of the well-known parametric forms are described by Davison et al. (2015) and Dey et al. (2016). An important part of these models is the formulation of the spatial covariance structure, which also defines the spatial extremal dependence. For a finite set of n points from $Z(\mathbf{s})$, spatial dependence is described by the extremal coefficient θ according to

$$\Pr[Z(\mathbf{s}_1) \leq z, \ldots, Z(\mathbf{s}_n) \leq z] = \exp(-\theta_n/z). \tag{21}$$

In Eq. 21, the extremal coefficient $1 \leq \theta_n \leq n$ such that the lower (upper) bound corresponds to perfect dependence (independence). Often, the bi-variate ($n = 2$) extremal coefficient is considered, as it is more practical to work with compared to higher dimensional cases. In practice, θ is typically estimated from the empirical counterpart obtained from data, for example using the F-madogram (Cooley et al., 2006).

The implementation of the max-stable process by Calafat et al. (2020) is based on the work by Reich et al. (2012), who provided a Bayesian description of the so-called residual max-stable process. The model is a finite approximation to the Gaussian extreme value process by Smith (1990) but with an added nugget term to account for the overly smooth spatial fields by the latter model and also to improve the computational stability. Noting the time-dependent annual maximum surge as $Y_t(\mathbf{s})$ and assuming that it is max-stable, the marginal distributions of observations were described as GEV$(\mu_t(\mathbf{s}), \sigma_t(\mathbf{s}), \xi)$. To simplify calculations, Calafat et al. (2020) assumed that the shape parameter stayed spatially constant. For the observations layer, the model likelihood stands as

$$Y_t(\mathbf{s}_i)|\theta_t(\mathbf{s}_i), \mu_t(\mathbf{s}_i), \sigma(\mathbf{s}_i), \xi, \alpha \overset{\text{ind}}{\sim} \text{GEV}(\mu_t^*(\mathbf{s}_i), \sigma_t^*(\mathbf{s}_i), \alpha\xi), \tag{22}$$

$$\mu_t^*(\mathbf{s}) = \mu_t(\mathbf{s}) + \frac{\sigma(\mathbf{s})}{\xi}(\theta_t(\mathbf{s})^\xi - 1), \tag{23}$$

$$\sigma^*(\mathbf{s}) \sim \alpha\sigma(\mathbf{s})\theta_t(\mathbf{s})^\xi. \tag{24}$$

Thus, it is assumed that the annual maxima at site $\mathbf{s}_i$ is independent conditional on the model parameters. Calafat et al. (2020) defined climatological dependence as spatial dependence in the GEV distribution parameters, which is captured by $\mu_t(\mathbf{s})$ and $\sigma(\mathbf{s})$. Residual dependence, which describes how individual sites are affected by the same storm, was modelled with the spatial residual process $\theta_t(\mathbf{s})$. The strength of the residual dependence is governed by $\alpha \in (0, 1)$ such that when $\alpha \to 0$ there is strong residual dependence and when $\alpha \to 1$ residual dependence vanishes.

The spatial residual process evaluated over L spatial knots is defined as

$$\theta_t(\mathbf{s}) = \left(\sum_{l=1}^{L} A_{t,l} w_l(\mathbf{s})^{1/\alpha} \right)^\alpha, \tag{25}$$

where $w_l(\mathbf{s}) \geq 0$ are scaled Gaussian kernel functions with temporally and spatially varying coefficient $A_{t,l}$ that follow a positive-stable distribution described in more detail by Stephenson (2003) and Reich et al. (2012). Using the positive-stable distribution ensures that the model is max-stable. To allow smooth variations in the location parameter, Calafat et al. (2020) used spatiotemporal integrated random-walk with Gaussian processes (Rasmussen et al., 2005). At time $t = 0$, $\mu_{t=0}(\mathbf{s}) \sim \text{GP}(\mathbf{x}^T(\mathbf{s})\boldsymbol{\beta}_\mu, c(\mathbf{s}, \mathbf{s}'; \gamma_{\mu_0}, \rho_{\mu_0}))$, where the design matrix $\mathbf{x}^T(\mathbf{s})$ contains the intercept and the width of the continental shelf at location $\mathbf{s}$ with regression coefficients $\boldsymbol{\beta}_\mu$. Note that this covariate is unlikely informative in shallow basins such as the Baltic Sea basin, and some other covariate more descriptive about the bathymetry could be more suitable. The covariance function c was modelled using the Matérn covariance function (Rasmussen et al., 2005), with the spread γ_{μ_0} and length scale ρ_{μ_0} estimated from the data. The trend term and the random-walk jump were modelled as zero-mean Gaussian processes with their own covariance structures. Similarly, the logarithm of the scale parameter $\log \sigma(\mathbf{s})$ was spatially modelled using Gaussian processes, but without temporal dependence. The rest of the details of the model can be found from Calafat et al. (2020).

The model was extensively validated with synthetic data generated from the model with known parameters, real-world observations and also using a storm surge reanalysis. All tests showed that the model generated adequate probabilistic reanalyses, with realistic uncertainty bounds and high correlation with the observed annual maxima. The return level values were more reliable than the single-site estimates, for which the estimated uncertainties were roughly twice as large as that of the hierarchical model. The results were less robust in some very localised cases, where the observations were starkly different between two neighbouring tide gauges. This could happen, when individual annual maxima are linked with very local storms and whose impact might be difficult to capture with the model. However, the GEV parameters were modelled with a high confidence to at least few hundred kilometers away from the nearest tide gauges, as they tend to vary rather smoothly across the domain.

A modified version of the same model was later used by Calafat et al. (2022), who estimated the contributions of internal climate variability and external forcing on storm surges in the same region as

Calafat et al. (2020). In their model, the scale parameter was kept constant and the contributions of internal climate variability and external forcing (through mean sea level changes) were incorporated in the integrated random-walk model for the location parameter. Their model was able to separate the spatial fingerprint of external forcing from the internal variability, which were then used to estimate the relative contributions of both factors to the spatial trends. They concluded that the internal climate variability can have a large opposite effect to the storm surge trends in comparison to external forcing, which underlines the need to account for the non-stationarity when calculating return level estimates.

Max-stable processes were also used by Rashid et al. (2024), who stochastically generated extreme value realisations from a max-stable process, using 41 tide gauges with 68 years of data along the coastal region of the United States. While the implementation details of the used max-stable process were not specified in the article, according to the code accompanied by the paper they fitted the Gaussian extreme value process model, the so-called Schlater's model (Schlather, 2002) with different covariance formulations and the Brown-Resnick model (Brown et al., 1977) with three regression formulations for the location and the scale parameter to the observations. Using the best fitting model, they generated a 10000-member ensemble of annual storm surge maxima along the coastal region of the United States, using annual maximum storm surges from the GTSM reanalysis. As their max-stable model generated samples that were temporally uncorrelated with the observations, spatiotemporal variability of the synthetic samples was matched with the observed annual maxima by re-ranking them according to the rank-ordering of the observed storm surge maxima. Ranks in ungauged locations were calculated as distance-weighted averages from the ranks of the two nearest tide gauges. This procedure was based on the assumption that the closest tide gauges provide sufficient information about the storm surge at a particular location of the coast, although for locations farther away from tide gauges, this assumption might not completely hold.

Validation of their approach showed that the synthetic annual maxima were slightly too large in many locations, although the error statistics were dominated by a single tide gauge in the Florida coast. Also, the uncertainty range in the estimated return levels was again substantially smaller compared to the single site fits. A major benefit of their approach is that it is feasible to generate a very large sample of sea level extremes, which allows to effectively quantify the uncertainty in the simulated values.

To conclude, fully spatial models provide the most comprehensive approach, along physics-based modelling, to assess sea level extremes in a spatially consistent way and can also be used to generate synthetic time series with appropriate data models. However, these methods can be computationally expensive and their use is currently limited in the large scale applications. Furthermore, their use is less straightforward in real-world studies compared to most of the simpler methods.

4.5 Satellite-based analyses

An alternative source of spatial information on sea level extremes is provided by global satellite altimetry. Two articles using satellite data were found that cover the study region. Lobeto et al. (2018), studied non-tidal residual storm surge on the US eastern coast using remote sensed sea level height from satellite altimetry. They fitted a non-stationary GEV distribution to the satellite observations and then corrected the spatial return level estimates $z_{r_i}^{\mathrm{SAT}}$ with an extreme scale factor according to the equation

$$z_{r_i}^{\mathrm{SAT}'} = \mathrm{ESF} z_{r_i}^{\mathrm{SAT}} = (0.65 K_{\mathrm{EP}} + 0.35 K_{\mathrm{W}}) z_{r_i}^{\mathrm{SAT}}. \tag{26}$$

In this equation, K_{EP} and K_{W} are parameters that describe the effects of exposure to the open ocean and the width of the continental shelf on the estimated return levels, respectively. The scaling factors were optimised using return level ratios with respect to the closest tide gauge $\phi_{r_i} = z_{r_i}^{\mathrm{SAT}'}/z_{r_i}^{\mathrm{TG}}$ that were defined according to the strength of correlation between the satellite and in-situ measurements. The scaling was designed to address errors, which are caused by missing satellite data and decreasing accuracy near the shore. After the scaling, the mean relative error was decreased substantially (from 86% to around 10%) in eight out-of-sample tide gauge locations. Their results showed that the developed methodology addressed some of the limitations of satellite altimetry data, which hamper their use in extreme value analysis and also helped to carry out spatiotemporal assessment of extreme sea levels.

Recently, Bij De Vaate et al. (2024) assessed the usability of global satellite-derived non-tidal residual storm surge for extreme value analysis, using 29 years of data from eight low-resolution mode satellite radar altimeters. They also used a non-stationary GEV model to cover full spatio-temporal variations in the sea level extremes. The satellite data was stacked on a 5° $\times$ 5° grid to increase the amount of data points in each grid box. Furthermore, a slightly different type of scaling compared to Lobeto et al. (2018) was applied to the GEV location and scale parameters in order to reduce the effect of under-sampling when estimating these two parameters. The scaling was performed using the GTSM ocean reanalysis instead of in-situ observations, as it was considered to better represent conditions over open ocean. The GEV distribution was fitted locally to 25 random sub-samples of the reanalysis time series after they had been trimmed to match the length of the local satellite time series. The scaling factors were then estimated by minimising the root-mean-squared-error between the GEV distributions fitted to the full and sub-sampled reanalysis time series with the median over the 25 samples used as the best estimate for the scaling factors. The results showed that the scaling factor for the location parameter was typically less than one, meaning that undersampling likely lead to overestimation of its values. In contrast, the scaling factor for the scale parameter showed less systematic variations throughout the global oceans. Bij De Vaate et al. (2024) concluded that while satellite altimetry data allows to estimate spatial variations in the global sea level extremes to certain extent, the coarse temporal resolution and the sensitivity to sea-ice make their use questionable at higher latitudes.

5 Conclusions

We have reviewed 37 articles, which considered some form of spatial extreme value analysis in the North Atlantic and Baltic Sea region. The reviewed articles were classified into five categories: 1) physics-based 2) regional frequency analysis, 3) one-dimensional, 4) fully spatial and 5) satellite-based extreme value analysis studies. The most common approach to spatial extreme value analysis was to run a physical model and then apply EVA to the simulated sea level. Models, by definition, provide physically consistent and spatially comprehensive data for analysis and are indispensable for climate change assessments of sea-level extremes. However, physical models are computationally heavy to run, and do not necessarily resolve local-scale features in storm surge behaviour. Furthermore, the results are to certain extent sensitive to atmospheric forcing used to drive the models. Regional frequency analysis and one-dimensional spatial models provide a simplified approaches to spatial modelling. The former pools observations over a certain region to increase the sample size, while the latter one typically applies statistical analysis over a coastal region, using the distance as an explanatory variable for the distribution parameters. These methods are usually simpler to apply than fully spatial models, but have limitations on how they capture the spatial dependency. Fully spatial models, on the other hand, particularly when applied to physical model data provide perhaps the most comprehensive approach to EVA, as they directly account for spatial dependence in sea-level extremes. Many methods also allow to generate spatial realisations of extremes from their data models. Although being rather complex and in some cases computationally heavy, these methods deserve a further inspection within MAWECLI.

One avenue worth further exploration is machine learning-based modelling of sea level extremes. While machine learning has been applied to storm surge prediction, there seems not to be many applications that also include EVA, particularly within the region of interest. An example of such a study was conducted by Rohmer et al. (2023), who used machine learning to increase the sample size when analysing wave height extremes in the Caribbean Sea region. It is expected that machine learning based modelling studies of extreme sea levels will gain popularity in the future.

6 Acknowledgments

We would like to thank M.Sc. Mikko Laapas for the internal review of the deliverable.

References

Andreevsky, M. et al. (2020). "Regional frequency analysis of extreme storm surges using the extremogram approach". en. In: *Natural Hazards and Earth System Sciences* 20.6, pp. 1705–1717. ISSN: 1684-9981. DOI: 10.5194/nhess-20-1705-2020.

Bardet, L. et al. (2011). "Regional frequency analysis of extreme storm surges along the French coast". en. In: *Natural Hazards and Earth System Sciences* 11.6, pp. 1627–1639. ISSN: 1684-9981. DOI: 10.5194/nhess-11-1627-2011.

Beck, N. et al. (2020). "Predicting extreme surges from sparse data using a copula-based hierarchical Bayesian spatial model". en. In: *Environmetrics* 31.5, e2616. ISSN: 1180-4009, 1099-095X. DOI: 10.1002/env.2616.

Bernardara, P. et al. (2011). "Application of regional frequency analysis to the estimation of extreme storm surges". en. In: *Journal of Geophysical Research* 116.C2, p. C02008. ISSN: 0148-0227. DOI: 10.1029/2010JC006229.

Bernier, N. B. et al. (2006). "Predicting the frequency of storm surges and extreme sea levels in the northwest Atlantic". en. In: *Journal of Geophysical Research: Oceans* 111.C10, 2005JC003168. ISSN: 0148-0227. DOI: 10.1029/2005JC003168.

Bij De Vaate, I. et al. (2024). "Mapping the spatiotemporal variability in global storm surge water levels using satellite radar altimetry". en. In: *Ocean Dynamics* 74.3, pp. 169–182. ISSN: 1616-7341, 1616-7228. DOI: 10.1007/s10236-023-01596-2.

Brown, B. M. et al. (1977). "Extreme values of independent stochastic processes". en. In: *Journal of Applied Probability* 14.4, pp. 732–739. ISSN: 0021-9002, 1475-6072. DOI: 10.2307/3213346.

Burchard, H. et al. (2002). *GETM, a General Estuarine Transport Model*. EUR Report EUR 20253 EN. JRC23237. European Commission.

Butler, A. et al. (2007a). "Extreme value analysis of decadal variations in storm surge elevations". en. In: *Journal of Marine Systems* 67.1-2, pp. 189–200. ISSN: 09247963. DOI: 10.1016/j.jmarsys.2006.10.006.

Butler, A. et al. (2007b). "Trend Estimation in Extremes of Synthetic North Sea Surges". en. In: *Journal of the Royal Statistical Society Series C: Applied Statistics* 56.4, pp. 395–414. ISSN: 0035-9254, 1467-9876. DOI: 10.1111/j.1467-9876.2007.00583.x.

Calafat, F. M. et al. (28, 2020). “Probabilistic reanalysis of storm surge extremes in Europe”. In: *Proceedings of the National Academy of Sciences* 117.4, pp. 1877–1883. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1913049117.

Calafat, F. M. et al. (31, 2022). “Trends in Europe storm surge extremes match the rate of sea-level rise”. In: *Nature* 603.7903, pp. 841–845. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-022-04426-5.

Coles, S. (2001). *An Introduction to statistical modeling of extreme values*. New York: Springer. DOI: 10.1007/978-1-4471-3675-0.

Coles, S. et al. (1990). “Statistics of coastal flood prevention”. en. In: *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences* 332.1627, pp. 457–476. ISSN: 0962-8428, 2054-0299. DOI: 10.1098/rsta.1990.0126.

Cooley, D. et al. (2006). “Variograms for spatial max-stable random fields”. en. In: *Dependence in Probability and Statistics*. Ed. by P. Bickel et al. Vol. 187. New York, NY: Springer New York, pp. 373–390. ISBN: 9780387317410 9780387360621. DOI: 10.1007/0-387-36062-X_17.

Cooley, D. et al. (2007). “Bayesian spatial modeling of extreme precipitation return levels”. In: *Journal of the American Statistical Association* 102.479. Publisher: Taylor & Francis, pp. 824–840.

Dalrymple, D. (1960). *Flood-frequency analyses, Manual of Hydrology: Part 3*. en. USGS Numbered Series, p. 80. DOI: 10.3133/wsp1543A.

Davison, A. et al. (2015). “Statistics of Extremes”. en. In: *Annual Review of Statistics and Its Application* 2.1, pp. 203–235. ISSN: 2326-8298, 2326-831X. DOI: 10.1146/annurev-statistics-010814-020133.

Deltares (2014). *Delft3D-Flow: Simulation of Multi-Dimensional Hydrodynamic Flows and Transport Phenomena, Including Sediments, User Manual*. Version 3.15.34158. 684 pp.

Dey, D. K. et al., eds. (6, 2016). *Extreme Value Modeling and Risk Analysis: Methods and Applications*. 0th ed. Chapman and Hall/CRC. ISBN: 9780429161193. DOI: 10.1201/b19721.

Dixon, M. J. et al. (1992). “Trends in UK extreme sea-levels: a spatial approach”. en. In: *Geophysical Journal International* 111.3, pp. 607–616. ISSN: 0956540X, 1365246X. DOI: 10.1111/j.1365-246X.1992.tb02115.x.

Dixon, M. J. et al. (1998). “Spatial modelling of extreme sea-levels”. In: *Environmetrics* 9.3, pp. 283–301. ISSN: 1180-4009, 1099-095X. DOI: 10.1002/(SICI)1099-095X(199805/06)9:3<283::AID-ENV304>3.0.CO;2-#.

Flather, R. A. et al. (1998). “First estimates of changes in extreme storm surge elevations due to the doubling of CO2”. In: *The Global Atmosphere and Ocean System* 6.2, pp. 193–208.

Frau, R. et al. (2018). “The use of historical information for regional frequency analysis of extreme skew surge”. en. In: *Natural Hazards and Earth System Sciences* 18.3, pp. 949–962. ISSN: 1684-9981. DOI: 10.5194/nhess-18-94a9-2018.

Gräwe, U. et al. (2012). “Storm surges in the Western Baltic Sea: the present and a possible future”. en. In: *Climate Dynamics* 39.1, pp. 165–183. ISSN: 1432-0894. DOI: 10.1007/s00382-011-1185-z.

Haan, L. D. (1984). “A Spectral Representation for Max-stable Processes”. In: *The Annals of Probability* 12.4. ISSN: 0091-1798. DOI: 10.1214/aop/1176993148.

Haigh, I. D. et al. (2010). “A comparison of the main methods for estimating probabilities of extreme still water levels”. en. In: *Coastal Engineering* 57.9, pp. 838–849. ISSN: 03783839. DOI: 10.1016/j.coastaleng.2010.04.002.

Heffernan, J. E. et al. (2004). “A Conditional Approach for Multivariate Extreme Values (with Discussion)”. en. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 66.3, pp. 497–546. ISSN: 1369-7412, 1467-9868. DOI: 10.1111/j.1467-9868.2004.02050.x.

Hosking, J. R. M. et al. (1997). *Regional Frequency Analysis: An Approach Based on L-Moments*. 1st ed. Cambridge University Press. ISBN: 9780521430456 9780521019408 9780511529443. DOI: 10.1017/CBO9780511529443.

Keef, C. et al. (2013a). "Estimating the probability of widespread flood events". en. In: *Environmetrics* 24.1, pp. 13–21. ISSN: 1180-4009, 1099-095X. DOI: 10.1002/env.2190.

Keef, C. et al. (2013b). "Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model". en. In: *Journal of Multivariate Analysis* 115, pp. 396–404. ISSN: 0047259X. DOI: 10.1016/j.jmva.2012.10.012.

Li, H. et al. (2023). "A spatially-dependent synthetic global dataset of extreme sea level events". en. In: *Weather and Climate Extremes* 41, p. 100596. ISSN: 22120947. DOI: 10.1016/j.wace.2023.100596.

Lobeto, H. et al. (2018). "Toward a Methodology for Estimating Coastal Extreme Sea Levels From Satellite Altimetry". en. In: *Journal of Geophysical Research: Oceans* 123.11, pp. 8284–8298. ISSN: 2169-9275, 2169-9291. DOI: 10.1029/2018JC014487.

Lorenz, M. et al. (2023). "Uncertainties and discrepancies in the representation of recent storm surges in a non-tidal semi-enclosed basin: a hindcast ensemble for the Baltic Sea". en. In: *Ocean Science* 19.6, pp. 1753–1771. ISSN: 1812-0792. DOI: 10.5194/os-19-1753-2023.

Lowe, J. A. et al. (2001). "Changes in the occurrence of storm surges around the United Kingdom under a future climate scenario using a dynamic storm surge model driven by the Hadley Centre climate models". In: *Climate Dynamics* 18.3-4, pp. 179–188. ISSN: 0930-7575, 1432-0894. DOI: 10.1007/s003820100163.

Lowe, J. et al. (2005). "The effects of climate change on storm surges around the United Kingdom". en. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 363.1831, pp. 1313–1328. ISSN: 1364-503X, 1471-2962. DOI: 10.1098/rsta.2005.1570.

Marcos, M. et al. (2012). "Effect of sea level extremes on the western Basque coast during the 21st century". en. In: *Climate Research* 51.3, pp. 237–248. ISSN: 0936-577X, 1616-1572. DOI: 10.3354/cr01069.

Marcos, M. et al. (2009). "Sea level extremes in southern Europe". en. In: *Journal of Geophysical Research: Oceans* 114.C1, 2008JC004912. ISSN: 0148-0227. DOI: 10.1029/2008JC004912.

Marcos, M. et al. (2011). "Changes in storm surges in southern Europe from a regional model under climate change scenarios". en. In: *Global and Planetary Change* 77.3-4, pp. 116–128. ISSN: 09218181. DOI: 10.1016/j.gloplacha.2011.04.002.

Medvedev, I. P. et al. (2016). "Tides in Three Enclosed Basins: The Baltic, Black, and Caspian Seas". In: *Frontiers in Marine Science* 3. ISSN: 2296-7745. DOI: 10.3389/fmars.2016.00046.

Meier, H. E. M. (2006). "Baltic Sea climate in the late twenty-first century: a dynamical downscaling approach using two global models and two emission scenarios". en. In: *Climate Dynamics* 27.1, pp. 39–68. ISSN: 0930-7575, 1432-0894. DOI: 10.1007/s00382-006-0124-x.

Meier, M. et al. (1999). *RCO – Rossby Centre regional Ocean climate model: model description (version 1.0) and first results from the hindcast period 1992/93*. Tech. rep. 26. MISU, p. 102.

Mellor, G. L. (1998). *Users Guide for a Three-Dimensional, Primitive Equation, Numerical Ocean Model*. Technical Report. Princeton, N.J.: Princeton University.

Mentaschi, L. et al. (2016). "The transformed-stationary approach: a generic and simplified methodology for non-stationary extreme value analysis". en. In: *Hydrology and Earth System Sciences* 20.9, pp. 3527–3547. ISSN: 1607-7938. DOI: 10.5194/hess-20-3527-2016.

Muis, S. et al. (2016). “A global reanalysis of storm surges and extreme sea levels”. en. In: *Nature Communications* 7.1, p. 11969. ISSN: 2041-1723. DOI: 10.1038/ncomms11969.
Muis, S. et al. (2017). “A comparison of two global datasets of extreme sea levels and resulting flood exposure”. en. In: *Earth’s Future* 5.4, pp. 379–392. ISSN: 2328-4277, 2328-4277. DOI: 10.1002/2016EF000430.
Muis, S. et al. (2019). “Spatiotemporal patterns of extreme sea levels along the western North-Atlantic coasts”. en. In: *Scientific Reports* 9.1, p. 3391. ISSN: 2045-2322. DOI: 10.1038/s41598-019-40157-w.
Muis, S. et al. (2020). “A High-Resolution Global Dataset of Extreme Sea Levels, Tides, and Storm Surges, Including Future Projections”. In: *Frontiers in Marine Science* 7, p. 263. ISSN: 2296-7745. DOI: 10.3389/fmars.2020.00263.
Muis, S. et al. (2023). “Global Projections of Storm Surges Using High-Resolution CMIP6 Climate Models”. en. In: *Earth’s Future* 11.9, e2023EF003479. ISSN: 2328-4277, 2328-4277. DOI: 10.1029/2023EF003479.
Pugh, D. et al. (1980). “APPLICATIONS OF THE JOINT PROBABILITY METHOD FOR EXTREME SEA LEVEL COMPUTATIONS.” en. In: *Proceedings of the Institution of Civil Engineers* 69.4, pp. 959–975. ISSN: 1753-7789. DOI: 10.1680/iicep.1980.2179.
Rashid, M. M. et al. (2024). “Stochastic simulation of storm surge extremes along the contiguous United States coastlines using the max-stable process”. en. In: *Communications Earth & Environment* 5.1, p. 39. ISSN: 2662-4435. DOI: 10.1038/s43247-024-01206-z.
Rasmussen, C. E. et al. (2005). *Gaussian Processes for Machine Learning*. en. The MIT Press. ISBN: 9780262256834. DOI: 10.7551/mitpress/3206.001.0001.
Ratsimandresy, A. et al. (2008). “A 44-year high-resolution ocean and atmospheric hindcast for the Mediterranean Basin developed within the HIPOCAS Project”. en. In: *Coastal Engineering* 55.11, pp. 827–842. ISSN: 03783839. DOI: 10.1016/j.coastaleng.2008.02.025.
Räty, O. et al. (2023). “Bayesian hierarchical modelling of sea-level extremes in the Finnish coastal region”. In: *Natural Hazards and Earth System Sciences* 23.7, pp. 2403–2418. DOI: 10.5194/nhess-23-2403-2023.
Reich, B. J. et al. (1, 2012). “A hierarchical max-stable spatial model for extreme precipitation”. In: *The Annals of Applied Statistics* 6.4. ISSN: 1932-6157. DOI: 10.1214/12-AOAS591.
Rohmer, J. et al. (2023). “Combining uncertain machine learning predictions and numerical simulation results for the extreme value analysis of cyclone-induced wave heights – Application in Guadeloupe”. en. In: *Ocean Modelling* 186, p. 102275. ISSN: 14635003. DOI: 10.1016/j.ocemod.2023.102275.
Rutgersson, A. et al. (2021). “Natural Hazards and Extreme Events in the Baltic Sea region”. In: *Earth System Dynamics Discussions* 2021, pp. 1–80. DOI: 10.5194/esd-2021-13.
Sang, H. et al. (2010). “Continuous Spatial Process Models for Spatial Extreme Values”. en. In: *Journal of Agricultural, Biological, and Environmental Statistics* 15.1, pp. 49–65. ISSN: 1085-7117, 1537-2693. DOI: 10.1007/s13253-009-0010-1.
Scarrott, C. et al. (2012). “Review of Extreme Value Threshold Estimation and Uncertainty Quantification”. en. In: *REVSTAT-Statistical Journal*, 33–60 Pages. DOI: 10.57805/REVSTAT.V10I1.110.
Schlather, M. (2002). “Models for Stationary Max-Stable Random Fields”. In: *Extremes* 5.1, pp. 33–44. ISSN: 13861999. DOI: 10.1023/A:1020977924878.
Shchepetkin, A. F. et al. (2005). “The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model”. en. In: *Ocean Modelling* 9.4, pp. 347–404. ISSN: 14635003. DOI: 10.1016/j.ocemod.2004.08.002.

Smith, R. L. (1990). “Max-stable processes and spatial extremes”. In: *Unpublished manuscript* 205, pp. 1–32.

Soomere, T. et al. (2018). “Variations in parameters of extreme value distributions of water level along the eastern Baltic Sea coast”. en. In: *Estuarine, Coastal and Shelf Science* 215, pp. 59–68. ISSN: 02727714. DOI: 10.1016/j.ecss.2018.10.010.

Stephenson, A. (2003). “Simulating Multivariate Extreme Value Distributions of Logistic Type”. In: *Extremes* 6.1, pp. 49–59. ISSN: 13861999. DOI: 10.1023/A:1026277229992.

Tawn, J. et al. (1989). “EXTREME SEA LEVELS; THE JOINT PROBABILITIES METHOD REVISITED AND REVISED.” en. In: *Proceedings of the Institution of Civil Engineers* 87.3, pp. 429–442. ISSN: 1753-7789. DOI: 10.1680/iicep.1989.2975.

Vousdoukas, M. I. et al. (2016). “Projections of extreme storm surge levels along Europe”. en. In: *Climate Dynamics* 47.9-10, pp. 3171–3190. ISSN: 0930-7575, 1432-0894. DOI: 10.1007/s00382-016-3019-5.

Vousdoukas, M. I. et al. (2018). “Global probabilistic projections of extreme sea levels show intensification of coastal flood hazard”. en. In: *Nature Communications* 9.1, p. 2360. ISSN: 2041-1723. DOI: 10.1038/s41467-018-04692-w.

Wahl, T. et al. (2017). “Understanding extreme sea levels for broad-scale coastal impact and adaptation analysis”. en. In: *Nature Communications* 8.1, p. 16075. ISSN: 2041-1723. DOI: 10.1038/ncomms16075.

Wang, S. et al. (2008). “The impact of climate change on storm surges over Irish waters”. en. In: *Ocean Modelling* 25.1-2, pp. 83–94. ISSN: 14635003. DOI: 10.1016/j.ocemod.2008.06.009.

Wang, S. et al. (2015). “Development and evaluation of a new regional coupled atmosphere–ocean model in the North Sea and Baltic Sea”. In: *Tellus A: Dynamic Meteorology and Oceanography* 67.1, p. 24284. ISSN: 1600-0870. DOI: 10.3402/tellusa.v67.24284.

Weiss, J. et al. (2013). “Comparison of local indices for regional frequency analysis with an application to extreme skew surges”. en. In: *Water Resources Research* 49.5, pp. 2940–2951. ISSN: 0043-1397, 1944-7973. DOI: 10.1002/wrcr.20225.

Weiss, J. et al. (2014a). “Formation of homogeneous regions for regional frequency analysis of extreme significant wave heights”. en. In: *Journal of Geophysical Research: Oceans* 119.5, pp. 2906–2922. ISSN: 21699275. DOI: 10.1002/2013JC009668.

— (2014b). “Modeling intersite dependence for regional frequency analysis of extreme marine events”. en. In: *Water Resources Research* 50.7, pp. 5926–5940. ISSN: 0043-1397, 1944-7973. DOI: 10.1002/2014WR015391.

Weisse, R. et al. (2021). “Sea level dynamics and coastal erosion in the Baltic Sea region”. In: *Earth System Dynamics* 12.3, pp. 871–898. DOI: 10.5194/esd-12-871-2021.