Modelling Extreme Water Level and its Components for Infrastructure Design

The case of Hook of Holland

Faidon Diakomopoulos

TU Delft

Modelling Extreme Water Level and its Components for Infrastructure Design

The case of Hook of Holland

by

Faidon Diakomopoulos

to obtain the degree of
Master of Science
in Civil Engineering
at the Delft University of Technology,
to be defended publicly on Monday November 13, 2023 at 11:45 AM.

Student number:	5621445	
Project duration:	March 1, 2023 – November 13, 2023	
Thesis committee:	Prof. Dr. Ir. E. Ragno,	TU Delft, supervisor
	Prof. Dr. Ir. A. Antonini,	TU Delft
	Prof. Dr. Ir. A. Bakker,	TU Delft/Rijkswaterstaat

Cover image source: Faidon Diakomopoulos

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

TU Delft

Summary

The projected increase in sea level is expected to increase the intensity of coastal flooding threatening communities living along the coast. This, in combination with population growth and urban expansion, calls for a better understanding of Extreme Water Levels (EWLs), the mechanisms generating them, and their components, i.e., astronomical tide and storm surge, since they drive the maintenance and design of flood protection systems. Netherlands' flood defense is crucial in facing the risk of flooding given its particular geographical configuration, its large number of inhabitants, and its high value of assets. For this, a better understanding of EWLs and their components is essential to assessing the quality of current structures and developing new adaptation strategies since they drive design and risk assessment procedures.

Here, in this paper, we compare different methods for estimating EWLs and their components (astronomical tides and surges) considering water level observations in Hook of Holland. More specifically, we present a step-wise procedure to investigate observed water levels and derive extreme conditions for design and risk assessment purposes. First, the extent to which observed water level is affected by sea surface pressure and wind speed, which are both considered physical drivers for storm surge generation, is investigated via spectral analysis, coherence function, and measure of dependence. Afterward, EWLs are extrapolated from observed water level, and their statistics, which are then used to infer design values beyond the range of observations, are derived following different approaches, i.e., univariate extreme value analysis on observed EWLs and two dependence models that explicitly account for EWLs components, copula functions, and the joint probability method.

The results show that storms in the Southwest Delta have a duration of about 4 days and that EWLs components, i.e., surge and astronomical tide, present negative dependence (the Kendall's tau $\tau = -0.50$). From the comparison between statistical approaches to model EWLs and infer design values, results show that copulas and JPM lead to an overestimation of EWL. However, EWLs modeled via copulas fit better low quantiles, as they imitate better the fluctuations of the real observations. Additionally the estimated EWL by the dependent copula for a probability of occurrence of $1/10000 years$ is equal to $423,3 cm$ quite close to the current literature ($\sim$ 420cm), while the estimation from the legal set of instruments for flood risk in the Netherlands (WBI 2017) varies from this ($\sim$ 510cm).

Keywords: Extreme Value Analysis, Copulas, Joint Probability Method (JPM), Tide-Surge interactions, Probabilistic Risk assessment

Thesis Structure

This MSc thesis is divided into two parts. Part I contains the main body of the thesis and it is written in a journal paper format since the aim is to submit it to the journal Earth System Dynamics (ESD). Part II constitutes supplementary material, including detailed analysis and additional information in support of the results in Part I. Both parts are intended to be published.

Contents

I

Paper

Abstract.

The projected increase in sea level is expected to increase the intensity of coastal flooding threatening communities living along the coast. This, in combination with population growth and urban expansion, calls for a better understanding of Extreme Water Levels (EWLs), the mechanisms generating them, and their components, i.e., astronomical tide and storm surge, since they drive the maintenance and design of flood protection systems. Netherlands' flood defense is crucial in facing the risk of flooding given its particular geographical configuration, its large number of inhabitants, and its high value of assets. For this, a better understanding of EWLs and their components is essential to assessing the quality of current structures and developing new adaptation strategies since they drive design and risk assessment procedures. Hence, in this paper, we investigate EWLs in Hook of Holland which represents a strategic location due to the inlet of the port of Rotterdam and the Maeslant storm surge barrier. Here, we present a stepwise procedure that starts by defining EWLs, assessing drivers of storm surges on observed sea levels via spectral analysis and coherence, and ends in estimating the statistics of EWLs based on multiple approaches, i.e., univariate extreme value analysis, copula functions, and Joint Probability Method (JPM). The results show that storms in the Southwest Delta have a duration of about 4 days and that EWLs components, i.e., surge and astronomical tide, present negative dependence (the Kendall's tau $\tau = -0.50$). From the comparison between statistical approaches to model EWLs and infer design values, results show that copulas and JPM lead to an overestimation of EWL. However, EWLs modeled via copulas fit better low quantiles.

1 Introduction

It is estimated that more than 1.2 billion people live in coastal regions within 100 km from the coastline (Small and Nicholls, 2003). Coastal cities and communities have been and will continue to be important economic and trade centers. However, population growth and urban expansion combined with projected Sea Level Rise (SLR) pose a significant threat to such communities(Nicholls, 1995; Woodruff et al., 2013; Neumann et al., 2015; Nicholls et al., 2008). Because of this, low-lying delta regions, where different natural processes interact, are under great pressure. For example, Hsiao et al. (2021) showed that the flooded area in Tawain due to coastal flooding can potentially increase between $17\% - 92\%$ in the future due to a changing climate. In Europe, the Ebro Delta region (Spain) requires a new flood management strategy to overcome the effect of projected SLR(Grases et al., 2020; Sánchez-Arcilla et al., 2008). In the US, the consequences of SLR in the Mississippi Delta were studied, and the importance of the integrated, long-term management plans was underlined (Day Jr and Templet, 1989). Generalizing the effect in low-lying delta regions on a global scale Nienhuis and Van de Wal (2021) estimated a loss of $\sim 5\%$ of global delta land in 2100 due to SLR.

The Netherlands is situated in the delta of the rivers Rhine, Meuse, Scheldt, and Ems along the North Sea coast. About 26 $\%$ of the Dutch territory is below mean sea level, and about 60 $\%$ is vulnerable to floods. Because of the large number of inhabitants and high value of assets, the Netherlands has a high level of flood protection, provided by a comprehensive system of dams, seawalls, storm surge barriers, dikes, dunes, pumps, sluices, and regular beach nourishments (Van Alphen et al., 2022). In this context, SLR is expected to alter extreme water level statistics, used for infrastructure design and risk assessment, are expected to change posing several issues regarding the suitability of the current management approach (Stijnen et al., 2014) and the operability of current protection systems (Van Alphen et al., 2022). In the Dutch Southwest Delta, a critical location is represented by Hook of Holland, where the inlet of the port of Rotterdam, the largest seaport in Europe, is located. Estimates of SLR for 2100 for this area range between $\sim$ 0.53 (SSP1-2.6) and $\sim$ 0.86m (SSP5-8.5), respectively (Garner et al., 2022). Here, the Maeslant storm surge barrier protects the Rotterdam harbor and the region of South Holland. The barrier was designed to close on average once every 10 years (Katsman et al., 2011; Van den Brink and de Goederen, 2017). However, it has been estimated that due to SLR the barrier might potentially close every 3 to 30 times per year at 1m and 1.5m SLR, respectively(Van Alphen et al., 2022). Hence, a deeper understanding of water level conditions, their components, and interactions, is essential to maintaining current flood safety levels and developing adaptation strategies in a changing climate (Antonini et al., 2019).

Extreme Water Levels (EWL) constitute the hydraulic loads for designing new infrastructure and assessing flood safety levels. Observed water level (WL) results from the combination of astronomical tides, driven by astronomical forcing, surge, influenced by weather systems interacting with the topography and morphology of the region, and mean sea level, i.e., the sea level when waves and tidal components are averaged out. To evaluate EWL, different approaches are implemented. In the UK, EWL is obtained considering high surges and high tides as potentially coincident (Horsburgh and Wilson, 2007). Another common design approach, more conservative, defines EWL as the sum of Highest Astronomical Tide (HAT) and maximum storm surge (Liu et al., 2010).

Many studies have shown the tendency of the peak of high surges to occur during rising tide (Proudman, 1955a, b; Rossiter, 1961). More recently (Arns et al., 2020) showed that the independence between storm surges and tides, often assumed for design purposes, can lead to an overestimation of EWL of up to 30%. Horsburgh and Wilson (2007) observed a phase difference between the peak of high surges and high tides in UK affecting observed EWLs. In the Adriatic and Tyrrhenian Sea, a negative relationship between high astronomical tide and surge is observed when investigating extreme water conditions (Ragno et al., 2023). In the Netherlands, Geerse et al. (2019) showed that there is a time difference between the peaks of high tides and the peaks of high storm surges and there is currently no evidence that they might coincide. On the other hand, it is worth noticing that Williams et al. (2016) showed the independence between extreme surges and astronomical tides when studying stations in UK, North Sea and the east coast of the US, meaning that any surge can coincide with any tide.

In the Netherlands, the probabilistic assessment tools used for EWL assessment on structures, e.g., the Maeslant barrier, evaluate WL as a combination of surges with a trapezoidal shape and tides with a cosine shape (Diermanse et al., 2013; Geerse et al., 2019; Geerse, 2020). For the calculation of EWL the phase difference between the peaks of storm surges and tides is not taken into consideration (Diermanse et al., 2013). After a revision for the legal set of instruments for flood risk (WBI 2017) a phase difference is introduced, i.e. $t_{peak,tide} - t_{peak,surge} = 4.5h$ (Geerse et al., 2019; Chbab, 2017) as the most frequently occurred during storms (Chbab, 2017), which results in a EWL equal to $\sim$ 510cm for a return period of 10000 years. Geerse et al. (2019) proposed a different phase difference between tides and surges, equal to -1.5h, with the extrapolated EWLs equal to$\sim$ 420cm for a period of 10000 years, which is in line with Van den Brink (2018). The latter estimations vary from the one given by the WBI 2017, hence a comparison and evaluation of these results with other statistical methods that have been developed is important for more accurate estimations of EWLs.

In this paper, we compare different methods for estimating EWLs and their components (astronomical tide and storm surge) considering water level observations in Hook of Holland. More specifically, we present a step-wise procedure to investigate observed water levels and derive extreme conditions for design and risk assessment purposes. First, the extent to which observed water level is affected by sea surface pressure and wind speed, which are both considered physical drivers for storm surge generation, is investigated via spectral analysis, coherence function, and measure of dependence. Afterward, EWLs are extrapolated from observed water level, and their statistics, which are then used to infer design values beyond the range of observations, are derived following different approaches, i.e., univariate extreme value analysis on observed EWLs and two dependence models which explicitly account for EWLs components, copula functions Arns et al. (2020); Ragno et al. (2023); Ferrarin et al. (2022) and the joint probability method Pugh and Vassie (1978). The main framework of our approach can be found in Figure 1.

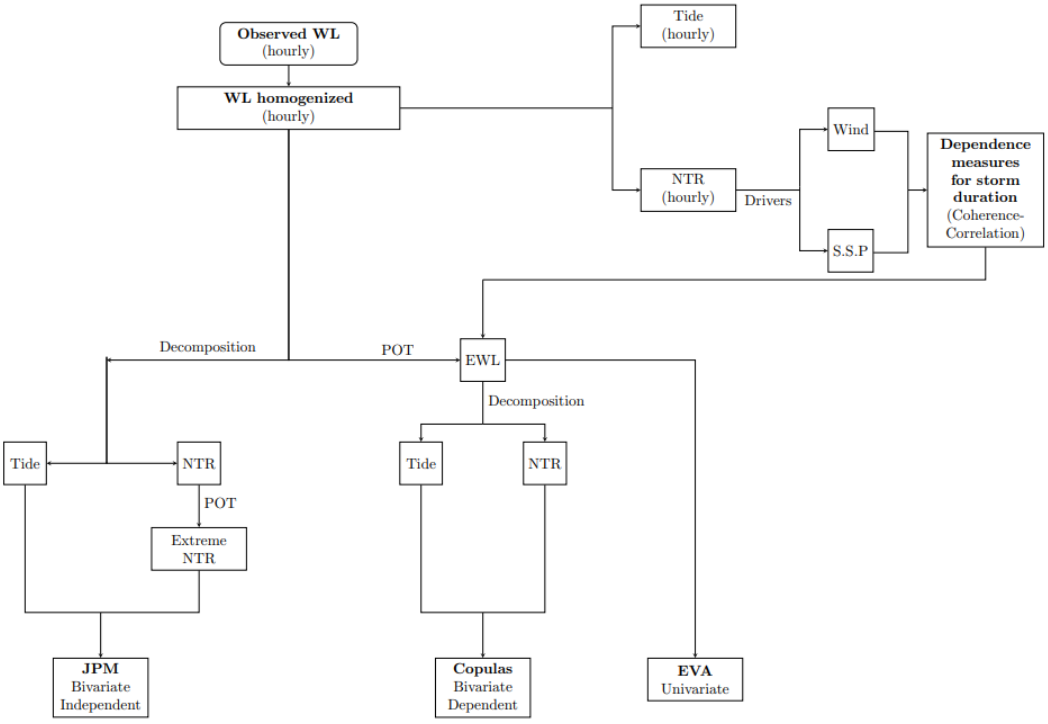


Figure 1. Flow diagram of the framework of the methodology adopted in the research paper.

The remainder of the paper is organized as follows. In Section 2. information about the location, and the importance of the study area, as well as the datasets are presented. The methods used to conduct the analyses are described in Section 3. Afterward, the results are presented in Section 4. Finally, the main points of discussion and conclusions are presented in Section 5 and Section 6, respectively.

2 Location and Data

We analyze observed Water Level (WL) in Hook of Holland (South Holland province, NL). This location is of great importance for the city of Rotterdam and port operations, as well as for the flood protection of the South Holland province due to the presence of the Maeslant storm surge barrier. Over the years, multiple man-made interventions especially targeted to the construction and expansion of the port, have altered the tidal range profile along the coast. For this, a pre-processing of WL observed will be necessary. Along with WL, sea surface pressure and wind speed and direction will be analyzed as well.

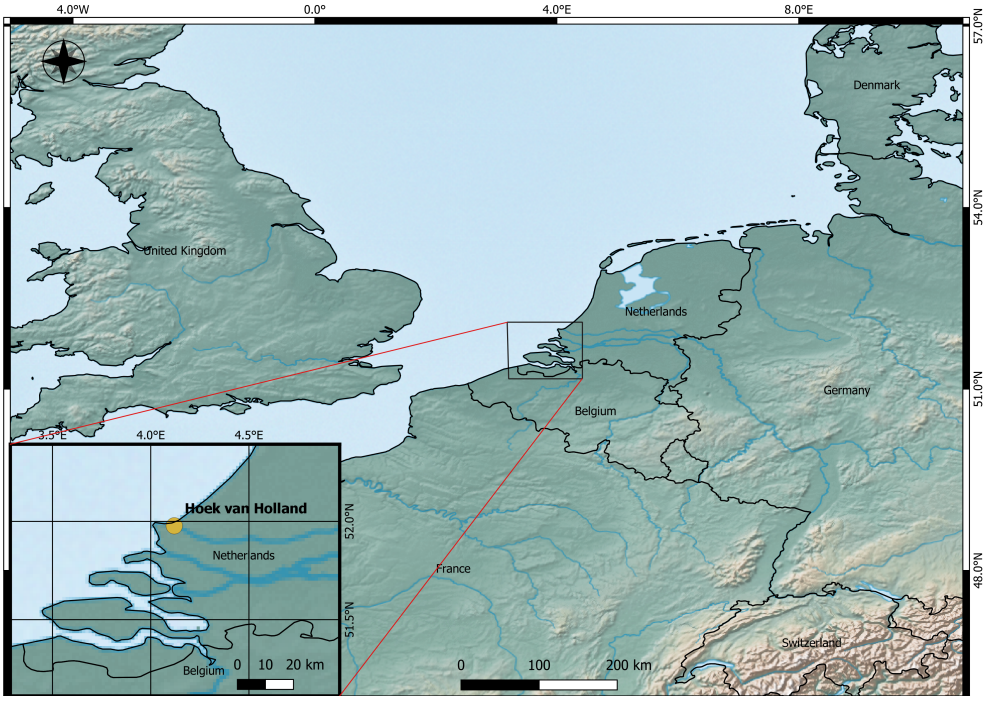


Figure 2. Hook of Holland Station.

2.1 Sea Level and its components

In this study, we analyze WL observations from Hook of Holland (gauge station located in 51°98.0'N°12.0'E and at a water depth of $\sim$ 10m). At this location, two different datasets are available, one from the Global Extreme Sea Level Analysis (GESLA) version 3 (Haigh et al., 2023) that contain sea water level observations on a global scale (see Sl Figure. A1 in the Supplement), and the other provided by Rijkswaterstaat (RWS), the executive agency of the Ministry of Infrastructure and Water Management in the Netherlands responsible for the management of water systems and waterways (Figure 3).

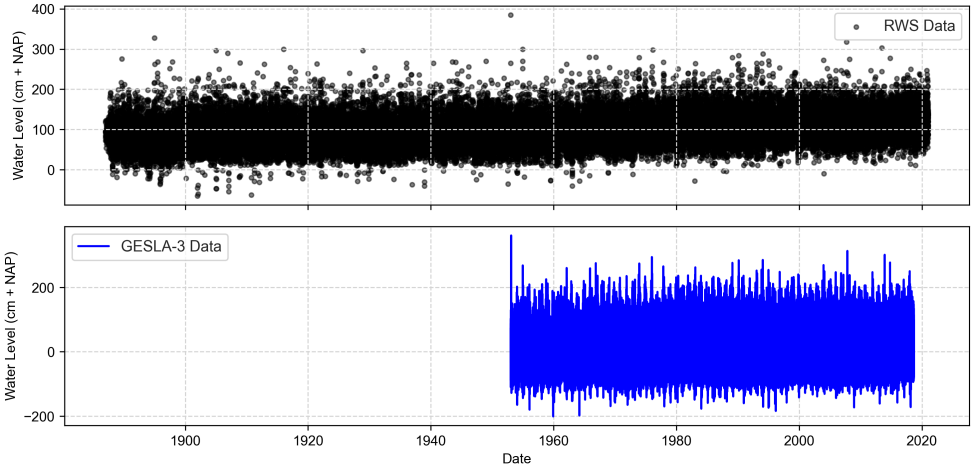


Figure 3. Comparison between RWS and GESLA-3 data for the Hook of Holland station. The maximum observed water level is captured in 01/02/1953 corresponds to the great flood of 1953 and the water reached up to 385 cm+NAP. The second highest peak (318 cm+NAP) that has been recorded in 09/11/2007 was extremely important as it forced to the closing of storm surge barrier for first time since its construction in 1997.

The two datasets differ in their temporal resolution and length of record. More specifically, RWS data contain two observations per day (highest peaks) for the period 1887-2020, while GESLA contains data with different hourly and sub-hourly frequencies for 1900-2018, respectively. GESLA data present gaps in the period 1900-1952, therefore these years are removed. Since our interest is in understanding WL and the processes generating EWLs, hourly observations are necessary. Hence, we select the GESLA dataset as the preferred dataset for the analyses. In support of this choice, we performed a comparative analysis between the two datasets based on which we could conclude that the statistics of GESLA dataset, which is quite shorter than RWS dataset, can sufficiently represent the statistics of RWS dataset. For more details on this preliminary analysis, the interested reader is referred to Supplementary Material (see Sl Figure 2).

The frequency of the GESLA dataset varies as follows: 1-hour frequency between 1953-1960, 3-hour frequency between 1961-1970, 30-min frequency between 1971-1986, and 10-min frequency between 1987-2018. In order to make the GESLA dataset uniform in terms of temporal resolution, the following procedure is applied: for sub-hourly observations, the maximum value recorded in one hourly is selected; for hourly observations, a linear interpolation between consecutive observations is performed as in (Pappas et al., 2014) assuming that only minor anomalies in the WL can occur between 1 and 3 hours.

Observed WL results from the combination of astronomical tides, surges from meteorological processes, and mean sea level when waves and tidal components are averaged out. In this study, we are interested in the tidal and surge components. Hence, the annual mean sea level is removed from hourly observations, accounting for man-made interventions. Details on how hourly WL are homogenized are in Section 3.1.

Astronomical tide, is considered the deterministic component of WL since they are connected with the position of the moon and the sun, although is affected by near-shore bathymetry. They are repeated on a daily scale and are reconstructed using the Matlab toolbox T_Tide by (Pawlowicz et al., 2002). The surge component, which is driven by weather patterns and the morphology of the area, is then evaluated by subtracting the reconstructed tide from WL observations. Because of this, we will be referred to the surge component as Non-Tidal Residuals (NTR) (Arns et al., 2020; Ferrarin et al., 2022).

2.2 Wind and Sea Surface Pressure

Sea surges in our area of interest are mainly caused by high winds from West and NNW directions (Diermanse et al., 2013; Groeneweg et al., 2022) that most of the time also occur during low atmospheric pressure systems. The latter phenomenon is known as the inverse barotropic effect (Weisse et al., 2012) in which the atmospheric pressure and the sea surface height are inversely proportional and the increase of atmospheric pressure leads to depressed WL and vice-versa. Therefore, we analyze wind speed (Figure 4) and sea surface pressure in Hook of Holland and made publicly available by the Dutch Meteorological Institute (KNMI) for the period between 1981 and 2018. The data do not present gaps and have a time resolution of 1 hour. All the datasets that have been used in this paper are tabulated and presented in Table 1.

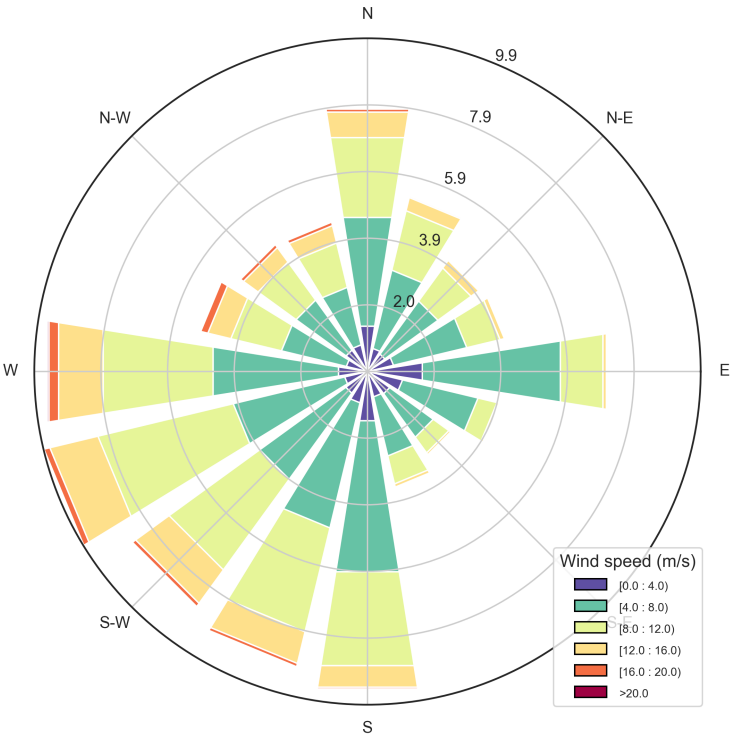


Figure 4. Wind rose for Hook of Holland station. The different colours express the wind speed classes while the numbers in the graph represent the frequency of cardinal directions in which the wind blows as percentages.

3 Methods

In this study, we present a step-wise procedure to analyse EWL and its components. Hence, multiple methods are implemented. Specifically, spectral and coherence analysis are implemented for analyzing hourly data, while probabilistic approaches, i.e., extreme value analysis, copula functions, and Joint Probability Method (JPM), are implemented for analyzing EWLs for design and risk assessment purposes. In this section, homogenization of WL observations and Peak Over Threshold (POT) approach

Table 1. Datasets of Water Level, Wind speed and direction, and Surface that have been used within this study.

Data set Name	Period	Time Resolution	Deleted Years	Total Number of Years
GESLA-3	1900-2018	1-hour (reassembled)	1900-1952	66
Rijkswaterstaat	1887-2020	2 peaks/day	-	133
KNMI-Wind Direction	1981-2018	1-hour	-	63
KNMI-Wind Speed	1981-2018	1-hour	-	63
KNMI-Surface Pressure	1981-2018	1-hour	-	63

for sampling extremes are also presented.

3.1 Data Homogenization and Trend's Test

The scope of this method is to homogenize the GESLA data by identifying and removing any trends and jumps, as it is a mandatory step to carry out the Extreme Value Analysis (Caires, 2011).

Due to the combination of many man-made interventions in Hook of Holland and evidence of SLR, homogenization of observed WL (detrending) is necessary in order to conduct statistical analysis as stationary data are needed. Here, Mean Sea Level (MSL) is calculated as the yearly average of hourly WL observations and it is removed from observations. The residuals are then adjusted to MSL observed in 2018. At the same time, we are interested in evaluating whether MSL presents long-term statistically significant trends. The most common method is the Mann-Kendall Test (Mann, 1945; Kendall, 1975) which tests the Null-Hypothesis (H_0) of no-trend in the data against alternatives. However, the correlation between data can lead to a faulty rejection of H_0 (Yue and Wang, 2004). To overcome this issue, the pre-whitening Mann-Kendall Test was introduced by (Yue and Wang, 2002) which adjusts the time series into independent events and removes serial correlation. Here, we implement the pre-whitening test at the significance level of $\alpha = 0.05$ to investigate the statistical significance of long-term trends on annual MSL.

3.2 Spectral Analysis

Spectral analysis is a widely used tool that decomposes a signal into simpler signals and evaluates the power spectra of those signals over different frequencies. High power spectra correspond to high amplitudes of the portion of the signal investigated, and vice-versa. Hence, spectral analysis of WL, NTR and tides enables us to identify the physical components leading to observed WL, e.g., the diurnal, semi-diurnal tides and non-linear effects, and their respective influence on the total signal (Simon, 2013; Medvedev et al., 2017, 2020; Schmitt et al., 2018).

Here, to determine components in WL, NTR and tides related to diurnal, semi-diurnal, and smaller periods of time, the spectrum is calculated using the Fast Fourier Transform (FFT) with different spectral resolutions. A spectral Kaiser Bessel window of $N = 8192$ h ($\Delta f = 1/N = 0.00122$ cph) with a half-window overlap and degrees of freedom $\nu = 81$ is used for identifying components with frequencies between $10^{-4} - 1$ cph (see also Thomson and Emery (2014); Medvedev et al. (2020)). Higher resolution spectral analysis, with a spectral Kaiser Bessel window of $N = 65536$ h ($\Delta f = 1/N = 0.000015$ cph) and a half-window overlap, allows us to observe interesting details focusing only in diurnal ($freq = 0.033 - 0.043$) and semidiurnal band ($freq = 0.077 - 0.085$) (Medvedev et al., 2017).

The low surface pressure is a critical factor for the generation of storm surges (Ferrarin et al., 2022; Woodworth et al., 2019), which are the stochastic part of the total water level. Hence, the spectral analysis of surface pressure is important for the

determination of frequencies with high energy peaks to understand possible periodicity and finally to investigate the coherence between water level and surface pressure. For this case, a spectral Kaiser Bessel window of $N = 65536$ h has been selected as it leads more clear results in the frequency domain.

3.3 Coherence

Coherence is a metric used to investigate the relationship between two signals (or timeseries) and assess the link between them. Coherence function is used to investigate the frequencies for which the two variables are correlated. Here, we use it to estimate declustering time of storms. More specifically, the declustering time is assumed to be the frequency corresponding to the highest coherence. The two variables for which the coherence is analyzed, NTR and sea surface pressure are correlated, because the low surface pressure systems (depression systems) cause storminess leading to NTR. When the highest coherence is reached this is an indicator that surface pressure and NTR are strongly correlated, hence the corresponding frequency is used as the window (declustering time) for independent events. The coherence function is given by the following equation:

$$C_{xy}(f) = \frac{|G_{xy}(f)|^2}{G_{xx}(f)G_{yy}(f)} \tag{1}$$

where, $G_{xx}(f)$ and $G_{yy}(f)$ are the power spectra of NTR and surface pressure, respectively, and $G_{\chi y}(f)$ is the cross-power spectrum between them. The range of $C_{xy}(f)$ varies from 1 which corresponds to perfectly related variables to 0 which means that there are no relationships between them. It denotes the simultaneous presence of energy peaks in the same frequency areas of the spectrum.

3.4 Statistical models for EWLs

We are interested in investigating EWLs and their statistics since they drive infrastructure design and risk assessment. Hence, we first describe the method adopted to select observations that can be considered representative of extreme events, i.e. Peak Over Thresholds approach. Then we describe the three different statistical models, i.e., extreme value analysis, copulas functions, and joint probability method, used to derive statistics of extremes and potential design values associated with low probability of occurrence.

3.4.1 Peak Over Threshold

Peak Over Threshold (POT) is a widely used approach to select peaks for modelling extreme events when dealing with sub-daily samples. POT method is a more sophisticated method based on on exceedances of a threshold compared to the annual maxima. The latter one method wasn't selected due to the sample size (66 years) because the estimates of the distributions based on this method would have larger variances (Caires, 2016).

Following the POT approach, a threshold u should be defined and excesses above such thresholds are considered extremes. The threshold should be high enough to ensure that only extremes are selected and simultaneously to ensure that a sufficient number of peaks are selected (Ashkar and Rousselle, 1987). Threshold selection can be a quite challenging issue, so many different formulations have been proposed. Graphical methods such as Mean Residual Life Plot (Davison and Smith, 1990) and Parameter Stability Plot can be used to select the appropriate threshold via visual inspection and as a result difficulties are encountered in finding the influence of small changes. Another option is connecting the number of observations (n), with the k-th highest water level $k = \sqrt{n}$ (Ferreira* et al., 2003) and $k = n^{2/3}/log[log(n)]$ (Scarrott and MacDonald, 2012). Another option, which is the most used is the selection of threshold value based on high percentiles, e.g., from 99th to 99.9th (Ferrarin et al., 2022; Arns et al., 2020; Wahl et al., 2017). It is worth noticing that this method depends on the record length and even if it is widely used it is difficult to be justified scientifically. Here, a threshold of 212.2 cm is chosen as it is in agreement with

the threshold of Dutch standards for EWL as a result of extreme storms from NNW direction (Diermanse et al., 2013). This value corresponds to the $\approx 99.9th$ percentile of the GESLA dataset used for the analysis. When the JPM is implemented, NTR peaks are selected via POT with a threshold of $140cm$, based on the stability parameters plot and mean residual life plot (see Sl Figure. D12 in the Supplement).

Once the threshold is selected, to infer the statistics of the extremes, such extremes should be independent, i.e., should come from independent events. To ensure the independence between storm events, a declustering time between two consecutive peaks should be selected. Usually, this declustering time coincides with the duration of a storm event. Previous literature about declustering time in Hook of Holland has reported 2.41 days as the time between two independent storm events (Dillingh et al., 1993), while more recent reports have seen that peaks with a time difference less than 4 days can be considered belonging to the same event (Caires, 2011). In this study, we use the results of the coherence analysis between NTR and surface pressure and the correlation between NTR and wind speed to select a representative declustering time.

3.4.2 Extreme Value Analysis

Extreme Value Analysis (EVA) is a method to determine the statistical characteristics of observed extreme events and extrapolate to low probability of occurrences that are often not observed, such as events occurring on average once every 1000 or 10000 years.

To select the distribution that best fits the observations, we investigate different distributions that can be categorized into two main families, according to their asymptotic behavior: "heavy tails" (or subexponential) and "light tails" (or superexponential or hyperexponential) distribution. Generally, "heavy tails" distributions are considered more promising for EVA compared to the light tails since the tail can contain more events (Papalexiou et al., 2013). Here, six distribution functions are investigated: the Generalized Pareto distribution (GPD) and Burr type XII (Burr, 1942) which belong to "heavy tail" or subexponential class; Generalized Gamma, Gamma, and truncated Gumbel with 3 parameters that belong to the "light tail" or superexpontial class; 4 parameters conditional Weibull distribution which belong to the superexpontial category as its shape parameter is larger than zero.

More specifically, the Probability Density Function (PDF) of three parameters Generalized Pareto distribution is given by:

$$f(x) = (1 + \xi \frac{x - \mu}{\sigma}))^{-1-\frac{1}{\xi}} \tag{2}$$

and it is defined for $x > \mu$, where μ is the location parameter equal to the threshold value from POT, $\xi \in \mathbb{R}$ is the shape parameter and σ is the scale parameter, affecting how heavy is the tail of the distribution (Papalexiou et al., 2013). It is used in previous analyses of extreme sea levels(Caruso and Marani, 2022; Wahl et al., 2017; Caires, 2011).

The Burr type XII distribution combines asymptotic properties of Weibull for low values and Pareto for the extremes (Koutsoyiannis et al., 2018). The PDF is given by:

$$f(x) = kc \frac{x^{c-1}}{((1 + (\frac{x-\mu}{\sigma})^c)^{k+1}} \tag{3}$$

It is defined for shape parameters $c, k > 0$, $x > \mu$, where μ is the location parameter equal to the threshold value from POT, and σ is the scale parameter.

The Generalized Gamma distribution can be defined as a generalization of Gamma, Weibull, and exponential distribution (Khodabina and Ahmadabadib, 2010) depending on the values of their parameters. Its PDF is given:

$$f(x) = \frac{k(\frac{x-\mu}{\sigma})^{ka-1} exp(-(\frac{x-\mu}{\sigma})^k)}{\Gamma(a)} \tag{4}$$

where $\Gamma(a)$ is the gamma function. It is defined for $x > \mu$, scale parameter $\sigma > 0$, and shape parameters $k, a > 0$.
The Gumbel distribution, which is widely used for sea water levels (Wahl et al., 2017), is given by:

$$f(x) = c\exp(\frac{x-\mu}{\sigma})\exp(-c(e^{\frac{x-\mu}{\sigma}} - 1)) \tag{5}$$

with shape parameter $c > 0$, $\sigma > 0$ as a scale parameter, while again $x > \mu$.

Gamma distribution can be seen as a sub-category of Generalized Gamma is given by:

$$f(x) = \frac{(\frac{x-\mu}{\sigma})^{a-1} exp(-(\frac{x-\mu}{\sigma}))}{\Gamma(a)} \tag{6}$$

The conditional Weibull distribution with 4 parameters is estimated as follows:

$$f(x) = a * c(1 - \exp(-(\frac{x-\mu}{\sigma})^k)^{a-1} \exp(-(\frac{x-\mu}{\sigma})^k)(\frac{x-\mu}{\sigma})^{k-1} \tag{7}$$

It is defined for shape parameters $k, a > 0$, while the μ and σ are the location and the scale parameter respectively.

It is worth noticing that when the available time-series are not long enough the estimation from this method could be unreliable (Caires, 2011).

For the parameters of all the aforementioned distributions, the Maximum Likelihood Estimation (MLE) method has been implemented. Then, two goodness-of-fit metrics are performed. The first metric is Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_i - x_i)^2} \tag{8}$$

where N is the total number of values, y_i and x_i are the predicted and observed values, respectively. A smaller number of this error indicates a better fit.

The other metric is the Kolmogorov-Smirnov test (K-S) which describes the maximum error in cumulative distribution functions:

$$K - S = max|F_{obs}(x) - F_{est}(x)| \tag{9}$$

where F_obs(x) and F_est(x) are the cumulative distribution of the observations and the fitted one, respectively. The test takes the largest absolute difference between the two distribution functions across all x values. The smaller the test statistic the better the fit.

These two goodness-of-fit metrics are used in extreme value analysis of wind speed (Dookie et al., 2018; Kollu et al., 2012). RMSE has been also used for EVA in waves (Naderi and Siadatmousavi, 2023). In this study, both of the tests are used for comparison and validation of "good" or "bad" fitting of the distributions to the data.

3.4.3 Dependence modelling: correlation and copulas

Extreme storm surges are driven by high wind speed. Hence, we quantify the dependence between NTR and wind speed by means of correlations, i.e., Spearman's ρ_s. Previous literature showed that along the Dutch coast, extreme storms are generated by westerly and northwesterly winds (Horsburgh and Wilson, 2007; Diermanse et al., 2013; Van den Brink and de Goederen, 2017). Following the approach presented in Lin et al. (2010), the critical direction of the wind speed component in the Netherlands is the direction between W and NNW which leads to the highest value of correlation. We calculate the correlation between NTR and wind speed considering different declustering times for independent storm events (i.e., 3, 4, 5 days) and different time lags between NTR peaks and wind speed peaks (i.e., from 12 up to 48 hours), as it is shown in Figure 5. In addition, we evaluate this correlation considering a subset of NTR and wind speed associated with the critical wind direction. In other words, we calculate the correlation following these steps: (1) Selection of all the NTR above 140cm (same as the threshold for JPM) for which in time t they correspond with the critical wind direction. After that, (2) different declustering times (3,4,5 days) were chosen to identify the time difference for 2 consecutive peaks of NTR. (3) Furthermore, a time window n adapted from 12 to hours equal to 36 hours inside of which the maximum value of wind speed was found before the peak NTR or after that corresponds to the critical direction. Finally, (4) the non-parametric Spearman's rank correlation coefficient (ρ) was calculated, to measure all the different correlations for declustering time between NTR peaks in days and time difference between maximum wind speed and NTR peak in a storm in hours. The equation of the coefficient is given below:

$$\rho = \frac{cov(R_X, R_Y)}{\sigma_{R(X)} * \sigma_{R(Y)}} \tag{10}$$

Where X and Y are the pair values for which the correlation is investigated, while R_X and R_Y are their ranks in ascending order. The values of the correlation coefficient range from -1 (negative correlation) to +1 (positive correlation). 0 means that the data are not correlated. The flexibility of Spearman's ρ coefficient can be used for every monotonic trend making it a better option for our dataset, in contrast with Pearson's ρ which is suitable only for linear trend.

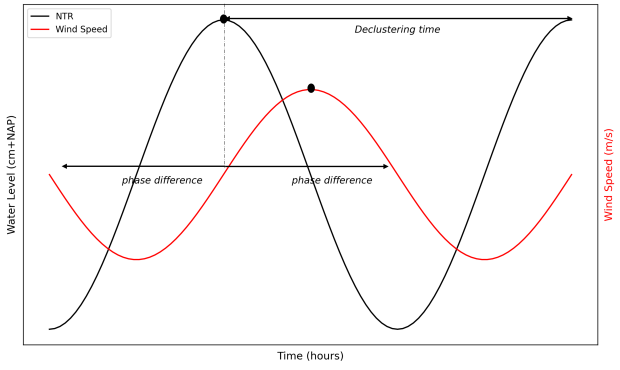


Figure 5. Explanation of the correlation method for NTR and wind speed. The declustering time determines the time difference between independent NTR peaks. The phase difference represents the time before and after the NTR peak during which the highest wind speed peak is correlated with the corresponding NTR peak (depicted by black dots).

For the copulas method, EWLs peaks derived from the POT analysis are decomposed into the corresponding NTR and tide components, as illustrated in Figure 6. The pairs of NTR and tide are used as follows.

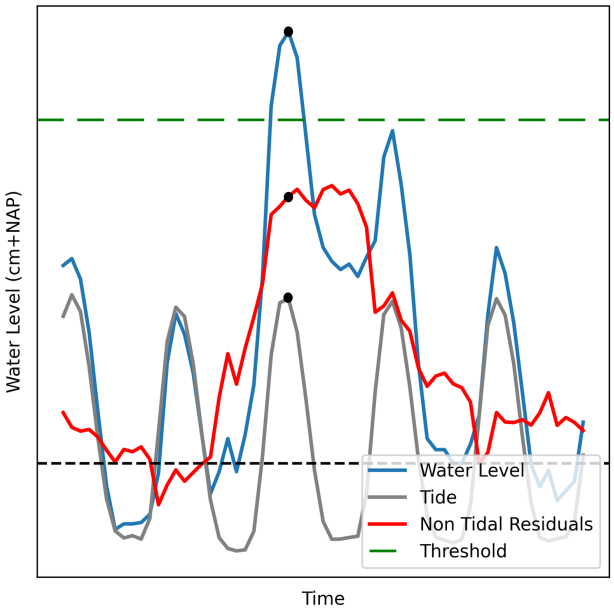


Figure 6. Explanation of the pair of NTR and tide that is chosen for the copulas method. The peak of the EWL exceeds the threshold of POT. The black dots in NTR and tide timeseries correspond to this peak of EWL at the same time t, so this is a chosen pair for the copulas method.

For the calculation of their dependence another metric for dependence, often used when dealing with copula functions is the non-parametric Kendall's rank correlation coefficient τ, (Kendall, 1938):

$$\tau = \frac{N_{cp} - N_{dp}}{N_p} \tag{11}$$

where N_{cp} is the number of concordant and N_{dp} the number of discordant pairs, respectively, and the sum of them $N_{cp} + N_{dp} = N_p$ the total number of pairs N_p. The pair of observations X, Y (in our case NTR and Tide) refers to (x_i, y_i) and (x_j, y_j) where $i < j$ and is defined as concordant if the sort order of (x_i, x_j) agrees with (y_i, y_j), otherwise it is called discordant. Its values vary between -1 for completely negative correlation, to $+1$ for positive, respectively. A value of 0 expresses that there is no ordinal correlation between observations.

If the two variables X, Y, with marginal distributions F_X and F_Y, show a level of dependence, their dependence structure can modeled via copula function C independently of their marginal distributions. Their joint distribution can be written :

$$F_{X,Y} = C(F_X(x)F_Y(y)) \tag{12}$$

where C is a unique copula (Sklar, 1973) when the marginal distributions are continuous. Following the approach of Ragno et al. (2023), the selection of the best copula is based on the Akaike Information Criterion (AIC).

$$AIC = 2(k - lnL) \tag{13}$$

where L is the likelihood function and k is the number of parameters of the model. The former can be seen as a penalization score for more complex copulas, i.e., with a higher number of parameters. The best copula is the one with the smallest AIC.

Along with the theoretical copulas, the independent copula is derived to assess the effect of modelling the dependence between NTR and tides on EWLs.

Copulas are modelled here using the Python library *pyvinecopulib* (https://github.com/vinecopulib/pyvinecopulib).

The comparison between the dependent and independent case is done by comparing EWLs randomly generated from the best theoretical copula and EWLs randomly generated from the independent copula. More specifically, 10000 pairs of dependent X_{dep}, Y_{dep} are sampled from the selected best copula, and 10000 pairs of independent X_{ind}, Y_{ind} are sampled from the independence copula. EWLs in both cases are obtained by the sum of the two components.

3.4.4 Joint Probability Method

Joint Probability Method (JPM), which is based on convolution, was first introduced by Pugh and Vassie (1978) and it is another method used for the estimation of EWL. For this method, peaks of NTR are selected and the underlying assumption is that NTR peaks are independent from tides.

The JPM estimates the distribution of EWL via the convolution of the Cumulative Distribution Function (CDF) of the NTR and the empirical Probability Density Function (EPDF) of all the tides.

$$F(z) = \int G(z - x) f(x) dx \tag{14}$$

in which F is the distribution of estimated EWL, G is the distribution of non-tidal components and f the empirical distribution of tides.

4 Results

4.1 Water Level Dynamics

This sub-chapter focuses on the storm events and how their correlation with the extreme non-tidal components. Additionally, using spectral analysis the main tidal constituents are highlighted. The coherence between water level and surface pressure is investigated, as well as the correlation between the wind speed and water level for the estimation of proper declustering time.

4.1.1 Mann Kendall Test trend analysis

The yearly mean water levels of GESLA data (Figure 7) show jumps connected with man-made interventions and upward trends from the rising of MSL. The data present no trend ($pvalue = 0.27$ for both tests) until the first breakpoint in 1965 when a jump probably associated with the works for Maasvlakte 1 and industrial area (Paalvast, 2014; Caires, 2011) is observed. For the period 1965-1990 the trend is strongly increasing $slope = 0.35$ ($pvalue = 0.0005$ for Mann-Kendall and $pvalue = p = 0.033$ for the pre-whitening test, respectively), followed by a sharp decline in 1990 caused by the connection Harteelkanaal- Beerkanaal. From the end of 1990 to 2020 the trend is still increasing ($pvalue = 0.027$ for Mann-Kendall and pre-whitening test), but with a milder slope than previously $slope = 0.1$, which could mean that the sea level rise is less evident. The trend analysis results of RWS data can be found on Sl Figure. B1 in the Supplement.

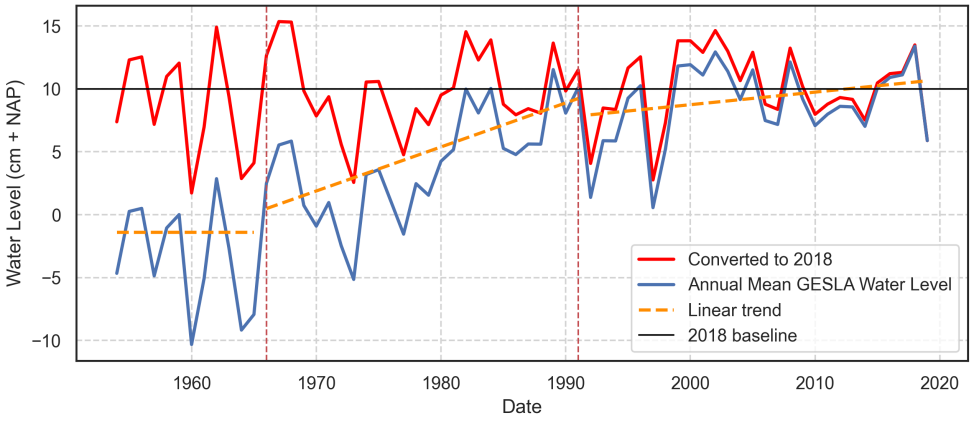


Figure 7. Illustration of the annual mean from GESLA data. The linear trends have been highlighted by the dashed orange line. The annual mean of homogenized and corrected data to the last year of observations (2018) are represented by the red line.

4.1.2 Spectral Analysis-Coherence

In the upper graph of Figure 8, the spectra of Water Level, NTR, and tide are shown. As the NTR is calculated by removing the tide from WL, there was expected negligible energy in the NTR spectrum in diurnal and semidiurnal bands as the WL in these bands is governed by the deterministic tides, so the Water Level energy and tidal energy there were expected to be almost the same. Nevertheless as can be seen from the graph energy peaks of NTR occur in these bands as a result of the interaction between NTR and tides. From the spectrum of NTR is shown that NTR still shows some level of tidal component as a result of the NTR-tide interaction.

In our station, the values of amplitude for the major components K_1, O_1, M_2 and S_2 are: 8, 11, 76, and 18cm, respectively with the value of F to be calculated as 0.2 corresponding to semi-diurnal dominance.

An interesting finding from the upper graph in Figure 8 is the energy peak to be reached in the corresponding from harmonic analysis component of fourth-diurnal M_4 ($freq = 0.17cph$) which is proportional to $(M_2)^2$ (Gräwe et al., 2014). This peak is the second highest exceeding the diurnal one. The high amplitude of M_4 component in the estuarine environment has been studied in different areas worldwide (Prestes et al., 2017; Guo et al., 2019) shown a high interaction between M_4 and M_2 components. Moreover, river discharge in the estuary seems to significantly influence the energy redistribution from principal tides to overtides (Guo et al., 2015). This could also explain the energy peak that corresponds with the M_6 $freq = 0.25cph$ overtide, whilst the peaks in higher frequency may be generated mainly by the influence of seiches in the area (Pattiaratchi,

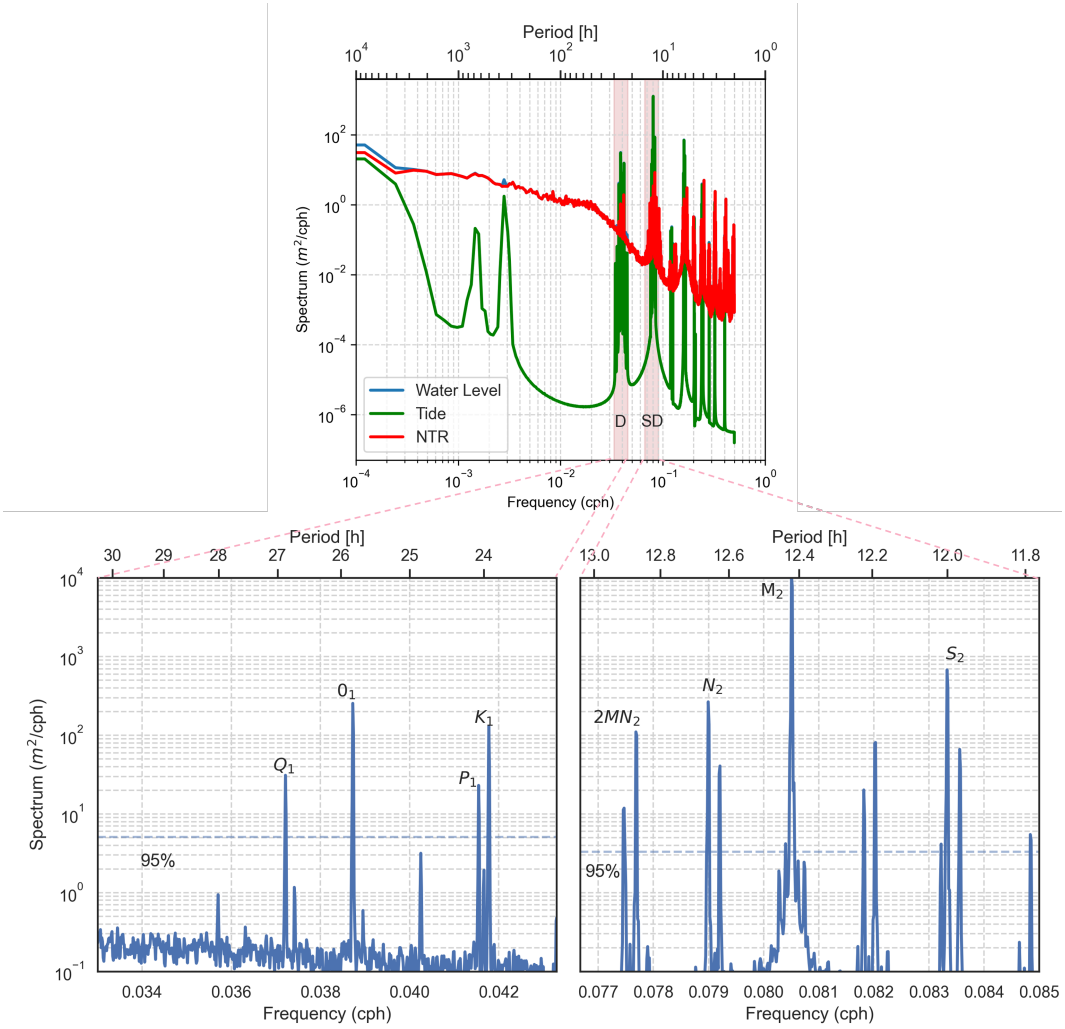


Figure 8. In the upper graph the spectral analysis for the total water level, NTR, and tides datasets is presented. The light red shaded areas represent the diurnal and semi-diurnal bands. As can be clearly seen from the upper graph there are energy peaks of the NTR spectrum in these bands because of the non-linear interaction between NTR and tides. It is obvious from the graph that the highest spectral energy is reached in the semi-diurnal region and especially in the M_2 constituent. This can be seen more clearly in the lower right graph which zooms in the semi-diurnal tidal components. The dominance of M_2, S_2, and N_2 constituents over the corresponding diurnal ones can be observed as their energies overpass the higher value that is connected in daily scale with the O_1 principal lunar component. In the lower graphs that focus on the diurnal and semi-diurnal areas the dashed line corresponds with the 95% confidence level and the most influenced harmonic constituents are highlighted (O_1, K_1, Q_1, P_1 and M_2, S_2, N_2, $2MN_2$, respectively).

2011). Another possible reason for the peaks of D_6 AND D_8 diurnal could be the interaction between surges, tides, and river water (Spicer et al., 2019).

From the two below graphs of the diurnal and semi-diurnal areas, it is obvious that M_2 is the predominant component in both regions, while O_1 is the dominant among the diurnal ones. In the diurnal area, it is evident that K_1 and P_1 that are connected

with the sun's declination are prominent, whereas the radiational S_1 component which is related to sea breezes is much smaller.

The major semi-diurnal constituents (M_2, S_2, N_2) are clearly distinguished, additionally, a significant peak is displayed in $2MN_2$ component as a result of the interaction of M_2 and N_2 due to frictional non-linearity (Teng et al., 2023). Around the major semi-diurnal peak (M_2), sharp peaks in the energy spectrum are observed. The known "tidal cusps" that are generated by the non-linear interaction between tides and mean sea level (Munk et al., 1965).

The spectrum of sea surface pressure has been analyzed presenting a predominance of semi-diurnal surface pressure variation (see Sl Figure. C1 in the Supplement), similar as in spectral analysis of tides. From the graph, it is difficult to observe peaks in smaller frequencies than in diurnal areas.

Regarding the results of coherence between NTR and sea surface pressure, the most interesting finding from Figure 9 is the frequency band ($0.01 - 0.02cph$) or in a period scale, from 50 to 96 hours, in which the coherence is slightly higher than 0.6 identifying the highest correlation between NTR and surface pressure. The period of 96 hours shows the link between surface pressure and NTR and can be used as an indicator of the window of the perturbation. Consequently, a declustering time of 4 days could be chosen for securing independence between storm events, as the highest coherence peak corresponds to 96 hours.

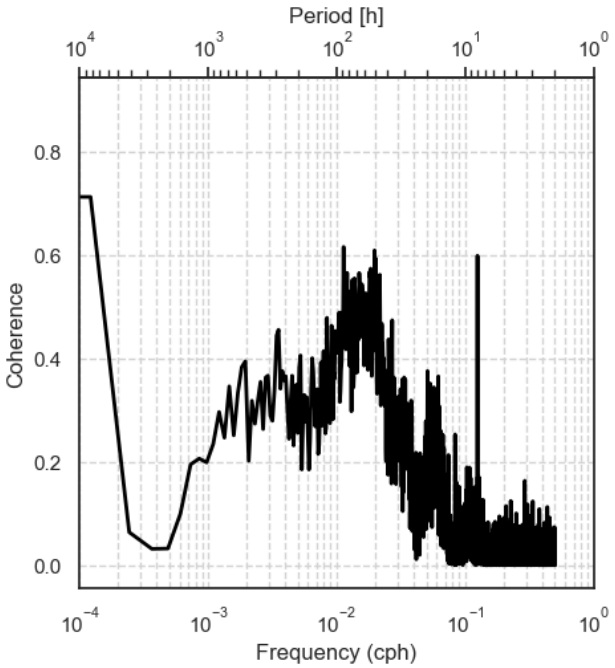


Figure 9. Coherence between Non Tidal Residuals (NTR) and surface pressure.

4.2 Water Level Extremes

4.2.1 Wind speed-NTR correlation

As it is mentioned the duration of the storm is crucial to identify and model EWL, as well as the wind direction. Hence we explore via measure of correlation the dependence between extreme NTR and wind speed during storm events.

The critical wind direction is found to be 340° as this leads to the highest correlation between NTR and wind speed when they correspond to a specific wind direction. The heatmaps of Figure 10,11 represent the values of Spearman's correlation coefficient for extreme NTR and wind speed. It is observed that all the correlations are positive as it was expected while the highest correlations when the wind speed component reaches the critical direction are more than two times higher compared to the correlation when the wind speed is independent of the wind direction (0.59 and 0.25, respectively). This could be an indicator of the importance of the direction from which the wind blows in the intensity of an extreme storm surge (Slomp et al., 2016). For the highest correlation on Figure 10 for a declustering time of 4 days and a time phase difference between extreme NTR and the peak of wind speed, the values are statistically significantly correlated ($pvalue = 0.05$). According to the declustering time, both graphs show similar correlations for 4 and 5 days, so the selection of 4 days can be justified by these results. According to the influence of the phase difference between peak NTR and peak wind speed in the critical direction of 340° the highest correlation is reached in 33 hours, while in Figure 11 this peak is observed between 15 and 18 hours.

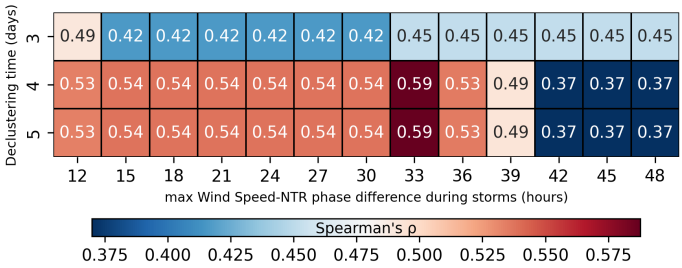


Figure 10. Spearman's correlation coefficient heatmap for Non Tidal Residuals and wind speed when the wind direction reaches the critical direction of 340°.

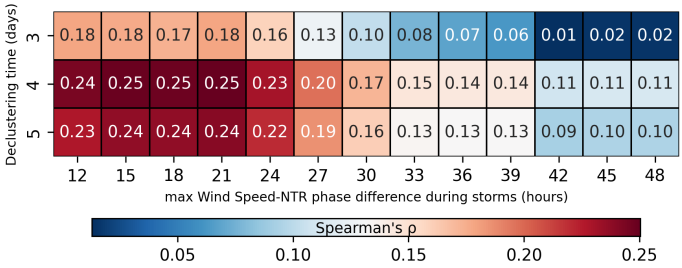


Figure 11. Spearman's correlation coefficient heatmap for Non Tidal Residuals and wind speed during storms independently of wind direction.

4.2.2 Storm Surge-Tide interaction

From the analysis of EWL above the threshold of $212.2cm$ the mean EWL is $236.6cm$ and the contribution of NTR is equal to $\approx 137cm$ and from the tidal one $99.6cm$, respectively. To express these values as percentages the effect of NTR is measured as 58% whilst for tides 42%. Moreover, for the 124 EWL from POT analysis in 99 of them the NTR is higher than the tide, and

only in 25 the situation reversed.

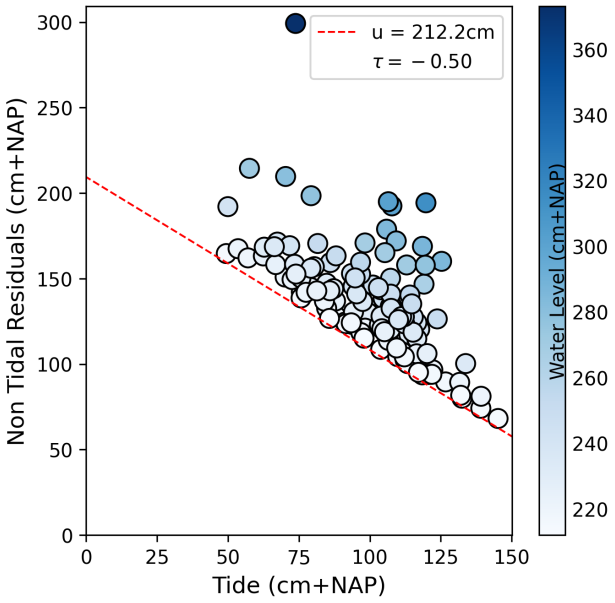


Figure 12. Scatter plot of NTR and tide interaction. The colorbar indicates the color of the dots connected with the EWL. As dots represent the sum of NTR and Tide the darkest ones correspond to the highest EWL, while the light to smaller value near to the EWL threshold of 212.2cm.

Apart of the relative contribution of NTR and tides, regarding to their dependency they are negatively dependent (Figure 12). The Kendall's tau correlation coefficient $\tau = -0.50$ shows a significant dependence $(pvalue = 0.02)$ between the values showing that the high NTR corresponds to low tides and the reverse. The results can be explained by the shallow water conditions on an estuary and the effect of the bed friction which leads to higher surge height in rising water and lower in high waters (Proudman, 1955a, b; Rossiter, 1961). These results have been validated also for the Adriatic sea with shallow water conditions (Ragno et al., 2023) and worldwide (Arns et al., 2020). From sensitivity analysis results we observed that the strength of the correlation coefficient reduces when the threshold is increasing. The reason for this can be that the dependence is connected in some way by the sampling approach in which the tides and NTR are conditioned by the EWL above threshold as it is expressed as their sum.

4.2.3 Statistical analysis

As the threshold has been determined to $212.2cm$ and the declustering time in 4 days the number of WL after POT method is 124. This means that for the 66 years of the given observations, the ratio of extreme events per year is $1.88event/year$. For all the distributions that have been evaluated in EVA, their graphs and estimation of EWL for the return period of $T = 10.000years$ can be found in Sl Figure. D7, D9 in the Supplement. The results of goodness of fit metrics for different distributions are given in Table 2. From these results it is shown that all the distributions fit well with the given GESLA data as their errors are quite small, so as a general comment every one of them could imitate the behavior of EWL sufficiently. However to identify the most accurate based on the best score of the tests (the smallest for RMSE and K-S), these are the Truncated Gumbel and Burr XII. The estimation of EWL for these two distributions is $432.2cm$ for Burr XII and $436.1cm$ for the Gumbel one, which are quite close to each other. On the other hand, Weibull distribution doesn't seem to fit so well with the GESLA data leading to

an overestimation of $EWL = 539cm$ with a probability of occurrence $1/10000$ years, the only distribution that produces an estimation of more than $500cm$.

Table 2. Values of RMSE and K-S for the distributions after Extreme Value Analysis.

Distribution Name	RMSE	K-S
Generalized Pareto	0.028	0.063
Burr XII	0.023	0.061
Generalized Gamma	0.024	0.062
Truncated Gumbel	0.023	0.061
Gamma	0.028	0.073
Conditional Weibull	0.034	0.075

In JPM, the POT method is applied on NTR leading to 141 events. The chosen distributions (see also Sl Figure. D13 of the Supplement) are Generalized Pareto with parameters shape parameter $\xi = 0.013$, location parameter $\mu = 140.42$ and scale $\sigma = 24.18$ and Burr XII with parameters $k = 1.162$, $c = 6.455$, location parameter $\mu = 140.43$ and scale $\sigma = 112.74$. The selection of these two distributions is based on their fitting on the NTR above threshold data, as the Burr XII due to its asymptotic behaviour estimates the NTR with a probability of occurrence 1/10000 years higher compared to all the other distributions whilst the corresponding value of NTR for this probability for Generalized Gamma is quite close to the other 4 distributions. The empirical Probability Density Function (PDF) of all the tides is illustrated in Sl Figure. D14 in the Supplement).

In copulas, the marginal distributions of the NTR and tide components need to be selected. In our case, the fit of distributions on data was poor (see Sl Figure. D15, D16 in the Supplement) so we decided to use empirical margins. We observed that the independence copula overestimates EWLs compared to the dependent copula, selected based on AIC is the Student. The Kendall's τ correlation coefficient from the NTR and tides that are generated by the copulas model is $\tau = -0.48$ quite close to the real correlation of $\tau = -0.50$.

As can be seen from Figure 13 the estimation of EWL, for the occurrence probability of $1/10000$ years is almost similar for both multivariate copulas model and univariate EVA. More specifically, the EWL estimated from the dependent Student copula is $423.3cm$ identical to the estimation of Burr XII from EVA, while the independent copula gives a higher estimation of $444.4cm$, slightly higher also from truncated Gumbel in EVA. Even if in Return Periods from 10 to 30 years the student copula slightly underestimates the EWL compared to the EVA, the main advantage of this method is that presents fluctuations, representing better the behaviour of observations compared to EVA which is a straight line in the graph. Especially the highest EWL observation which corresponds to the flood of 1953, we see how influence the dependent copula to present a peak in this Return Period, imitating the peak in the observations. Additionally, the oscillations in higher Return Periods can be results of the interaction between surges and tides, leading to a most realistic result. The behaviour of Empirical copula (TLL) is comparable with the dependent copula, and especially in Return Periods smaller than 30 years fits almost perfectly to the data, but in 10000 years seems to underestimate the EWL. On the contrary, the independent copula systematically overestimates the EWL presenting fluctuations as the Student one, but deviates significantly from the observed data. According to the JPM method, doesn't fit sufficiently in the GESLA data, as in Return Periods from 3 to 100 years underestimates the WL whilst in the specified return period of 10000 years both of the distributions of the method overestimate the EWL. Generalized Pareto JPM leads to an EWL of $462cm$ and Burr XII in $534cm$, respectively.

The difference in the results of copulas and JPM can possibly be explained as the two approaches differ in the principles of EWL. To be more precise, dependence copulas-based modelling is EWL driven, as the pair values of NTR and tide that correspond to a WL above the threshold of 212.2 cm are chosen. Consequently, in the calculation of dependence and generation of dependent samples, their sum will be similar to EWLs from observations. On the contrary in JPM method is driven by the peaks of NTR above threshold. The NTR-tide independence means that in every peak NTR a value of tide is added, leading to

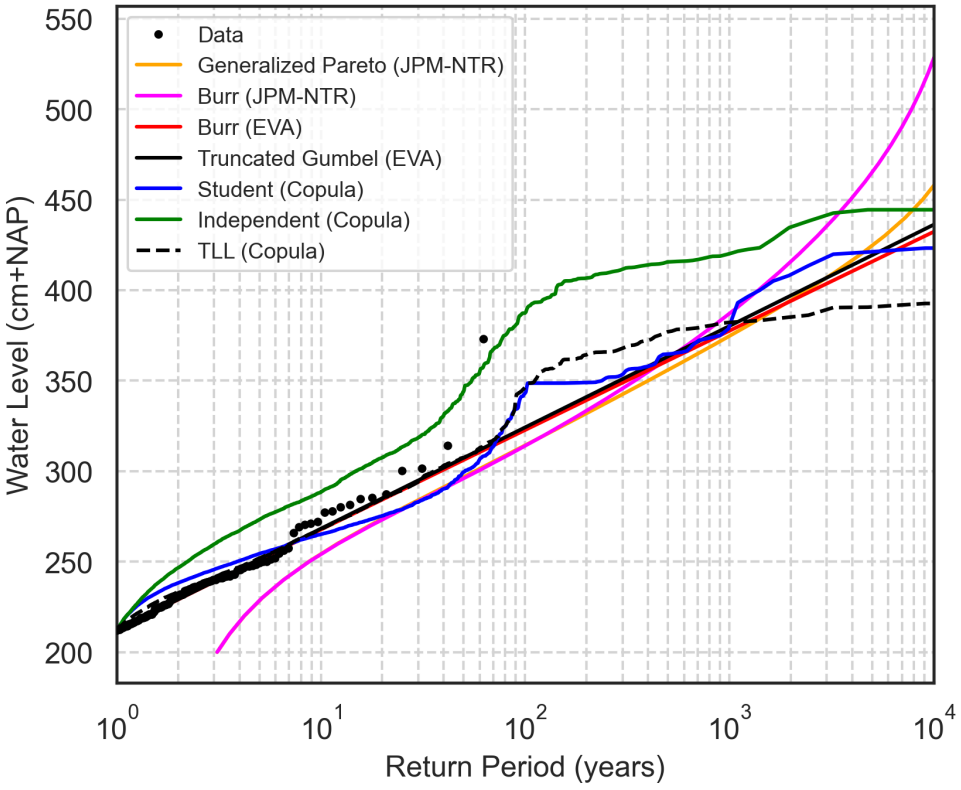


Figure 13. Illustration of the results of different statistical methods for the estimation of EWL.

an estimation of EWL that is not related, a priori with the observed EWLs.

5 Discussion

The current practices for the estimation of EWL seem to have some weaknesses according to the worst-case scenario as they assume a coincidence of extreme NTR and high tides that generate the EWL. We analyzed the dependence between surges and tides that their sum is leading to EWL. The analysis is focused on the location of Hook of Holland. The results are referenced to an estuarine environment with the effect of shallow water conditions and bottom friction and cannot be generalized for every sea level condition. Nevertheless, this does not affect the validity of the results of the estimation of EWL. From these results, it is shown that the hypothesis that EWL is a combination of high tides and extreme surges can lead to an overestimation of EWL for higher Return Periods. The results of JPM method and independent copulas in which extreme NTR can coincide with high tides, estimate EWL higher than the univariate EVA. On the other hand, taking into consideration the dependence between tides and NTR it is possible to improve the variability of the estimation of EWL.

In POT method the influence of threshold value and declustering time in the total number of Extreme Water Levels is evaluated (see Sl Figure. D3, D4 in the Supplement). From the results can be seen that the threshold value contributes decisively in the number of events whilst the declustering time is less influential. Sensitivity analysis for the influence of threshold on EVA has also been implemented (see Sl Figure. D11 in the Supplement). From the results, it seems that as the threshold increases

Burr XII and Generalized Gamma present the best scores in goodness of fit metrics. For the results of the estimation of EWL with a probability of occurrence 1 in 10000 years Burr XII distribution leads to higher values as the threshold increases. These results can be an indicator of the behaviour of distributions for different thresholds, but the visual inspection of the distribution fitting for the specific threshold is of interest.

According to the influence of wind speed to extreme NTR the statistical dependence is calculated. It is shown that including the critical direction as proposed in previous literature (Lin et al., 2010; Groeneweg et al., 2022) the correlation seems to be increased. For Hook of Holland, the number of NTR that are above the threshold of 140cm and correspond with the critical direction is 38 in the period 1981-2018, before being divided into independent events by the declustering time. The small number of events however doesn't affect the validity of the result since the a p value test is implemented. This correlation doesn't seem to be critically influenced by the declustering time of NTR in the range of 2 to 4 days. However, it is affected by the time difference window between peak NTR and peak wind speed. The value of the critical wind speed to maximize the correlation ($340°$) is in agreement with previous findings about the storms in the region that are generated by extreme westerly and northwesterly winds (Horsburgh and Wilson, 2007; Van den Brink and de Goederen, 2017), while for the dutch flood management by storms the most extreme surges are correlated with NNW direction (Diermanse et al., 2013). Moreover as can be seen from the sensitivity analysis about the threshold of NTR and the correlation it is shown that the highest threshold leads to the highest correlation (see Sl Figure. C2).

Copulas modelling is implemented to generate larger samples of the pair NTR and tides accounting for their dependence structure. Therefore events can be extrapolated up to probabilities of occurrence of $1/1000$ and $1/10000$. The selection of a theoretical copula is significantly important since it can lead to different behaviors, especially in the tails, as shown in Sl Figure.D17 in the Supplement. Additionally, for the marginal distributions of the NTR and tides a poor fitting of their respective theoretical distributions can lead to unreliable estimations. In such a case, the empirical distributions are preferred.

NTR shows some level of deterministic component that is connected with the deterministic part (tides) in the diurnal and semidiurnal bands. As a result is questionable if the selected peaks for JPM are independent because as it is shown from spectral analysis there is a connection between them and the deterministic tides, as a result of the interaction between them.

The main difference between dependence copulas and JPM is that copulas are related to the EWLs whilst JPM is related to NTR above threshold. As a result, copulas can estimate the EWLs closer to real observations than JPM which is based on the threshold selection of NTR, and the addition of any tide in an extreme NTR, leading to an estimation of EWL, when in the JPM method the dependence of the pair of NTR and tide that corresponds to an EWL is not taken into consideration.

6 Conclusions

We analyzed the EWLs in Hook of Holland both from the perspective of water level dynamics as well as of water level extremes. We identified the coherence between WL and sea surface pressure, finding an indicator of the assumption of declustering time equal to 4 days in the high coherence that is achieved for this period. After the investigation of the correlation between extreme surges and peaks in wind speed data, we investigated the effect of the critical direction of the wind speed component as well as we validated the selection of declustering time from the coherence analysis. Including the critical direction the calculated correlation is higher which can explain the importance of this direction in the storm surges.

Additionally, by the spectral analysis, we tried to better understand the complicated estuarine environment of Hook of Holland station with the interaction of tides and specific conditions to lead in high amplitudes of D4, D6, and D8 components. Moreover, the man-made interventions that changed the profile of mean sea level are referenced as well as the influence of sea level rise until the end of our data is taken into consideration.

Another important issue of this work was to quantify the surge-tide interaction which especially in shallow water conditions and coastal regions plays an important role in the estimation of EWL. The result of the negative correlation between NTR and tides in combination with the much bigger influence of NTR in the EWL, compared to tides shows that the use of statistical models for EWL, that can taken into consideration this interaction can improve the estimation of EWLs for this area. Keeping this in mind we proposed the use of copula copula-based model based on the dependence of NTR and Tide. EVA results were utilized as a reference point and the independent JPM model is compared. The results of this analysis show:

- From EVA analysis results for Hook of Holland a subsample of 66 years such as GESLA data, represents sufficiently the longer datasets from 1889 of RWS.
- The independence assumption can lead to misleading of the EWLs with a low probability of occurrence such as $1/10000$ years which was the design standard for Maeslant barrier. Both the JPM method as well as the independent copula tend to overestimate the EWL. For the former one in the case of Burr distribution, the estimation of EWL was almost one meter higher than the estimation of EVA and copulas, which could cause a significant increase in the design and maintenance cost of structures.
- In the EVA method the conditional Weibull distribution leads to higher estimations of EWLs compared to all the other distributions within this study for the given dataset of 212.2cm. As the EVA is really sensitive to the threshold selection and the number of data, these results are compared with RWS data for the same threshold, leading again to errors higher than the others.
- The estimated EWL on a return period by the student copula is $423.3cm$ which is quite close to the results from current literature for Hook of Holland that estimate EWL equal to $\sim 420cm$ for the same return period (Geerse et al., 2019; Van den Brink, 2018). On the other hand, WBI 2017 estimation of $\sim 510cm$ varies from all the estimated EWLs apart of the JPM method with Burr XII distribution ($434cm$).
- Additionally, copulas seem to overestimate and underestimate WL in low Return Periods, but dependent copula prevails as imitates better the real observations presenting fluctuations based on the values of the given dataset for low quantiles in which this study is focusing.

Data availability. The data used in this study are publicly available and can be retrieved from "Rijkswaterstaat" (Executive Agency of the Ministry of Infrastructure and Water Management in the Netherlands) https://www.rijkswaterstaat.nl/ and Global Extreme Sea Level Analysis (GESLA) https://gesla787883612.wordpress.com/ for water level observations. For wind and surface data, they can be downloaded from "Koninklijk Nederlands Meteorologisch Instituut" (KNMI) https://www.knmi.nl/home

References

Antonini, A., Raby, A., Brownjohn, J. M. W., Pappas, A., and D'Ayala, D.: Survivability assessment of fastnet lighthouse, Coastal Engineering, 150, 18–38, https://doi.org/https://doi.org/10.1016/j.coastaleng.2019.03.007, 2019.

Arns, A., Wahl, T., Wolff, C., Vafeidis, A. T., Haigh, I. D., Woodworth, P., Niehüser, S., and Jensen, J.: Non-linear interaction modulates global extreme sea levels, coastal flood exposure, and impacts, Nature communications, 11, 1918, https://doi.org/https://doi.org/10.1038/s41467-020-15752-5, 2020.

Ashkar, F. and Rousselle, J.: Partial duration series modeling under the assumption of a Poissonian flood count, Journal of Hydrology, 90, 135–144, https://doi.org/https://doi.org/10.1016/0022-1694(87)90176-4, 1987.

Burr, I. W.: Cumulative frequency functions, The Annals of mathematical statistics, 13, 215–232, https://doi.org/https://www.jstor.org/stable/2235756, 1942.

Caires, S.: Extreme value analysis: Still water level, Geneva, Switzerland: World Meteorological Organization, Intergovernmental Oceanographic Commission (of UNESCO), JCOMM Technical Report, 2011.

Caires, S.: A comparative simulation study of the annual maxima and the peaks-over-threshold methods, Journal of Offshore Mechanics and Arctic Engineering, 138, 051 601, https://doi.org/https://doi.org/10.1115/1.4033563, 2016.

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

Chbab, E.: Basisstochasten WBI-2017- Statistiek en statistische onzekerheid, Deltares Report, 2017.

Davison, A. C. and Smith, R. L.: Models for exceedances over high thresholds, Journal of the Royal Statistical Society Series B: Statistical Methodology, 52, 393–425, https://doi.org/https://www.jstor.org/stable/2345667, 1990.

Day Jr, J. W. and Templet, P.: Consequences of sea level rise: implications from the Mississippi Delta, Coastal Management, 17, 241–257, https://doi.org/https://doi.org/10.1080/08920758909362088, 1989.

Diermanse, F., Roscoe, K., de la Cruz, J. L., Steenbergen, H., and Vrouwenvelder, T.: Hydra Ring Scientific Documentation, Deltares Report, 2013.

Dillingh, D., De Haan, L., Helmers, R., Können, G., and Van Malde, J.: De basispeilen langs de Nederlandse kust: Statistisch onderzoek, DGW-93.023 rapport Dienst Getijdewateren, 1993.

Dookie, I., Rocke, S., Singh, A., and Ramlal, C. J.: Evaluating wind speed probability distribution models with a novel goodness of fit metric: a Trinidad and Tobago case study, International journal of energy and environmental engineering, 9, 323–339, https://doi.org/https://doi.org/10.1007/s40095-018-0271-y, 2018.

Ferrarin, C., Lionello, P., Orlić, M., Raicich, F., and Salvadori, G.: Venice as a paradigm of coastal flooding under multiple compound drivers, Scientific reports, 12, 5754, https://doi.org/https://doi.org/10.1038/s41598-022-09652-5, 2022.

Ferreira*, A., de Haan*, L., and Peng¶, L.: On optimising the estimation of high quantiles of a probability distribution, Statistics, 37, 401–434, https://doi.org/10.1080/0233188021000055345, 2003.

Garner, G., Hermans, T. H., Kopp, R., Slangen, A., Edwards, T., Levermann, A., Nowicki, S., Palmer, M. D., Smith, C., Fox-Kemper, B., et al.: IPCC AR6 WGI Sea Level Projections, https://doi.org/10.1017/9781009157896.011, 2022.

Geerse, C.: Interaction between tide and surge-Modelling the tide evolution of the residual surge, HKV consultants Final Report, 2020.

Geerse, C., Rongen, G., and Strijker, B.: Schematization of storm surges-Analysis based on simulated KNMI-data, HKV consultants Final Report, 2019.

Grases, A., Gracia, V., García-León, M., Lin-Ye, J., and Sierra, J. P.: Coastal flooding and erosion under a changing climate: implications at a low-lying coast (Ebro Delta), Water, 12, 346, https://doi.org/https://doi.org/10.3390/w12020346, 2020.

Gräwe, U., Burchard, H., Müller, M., and Schuttelaars, H. M.: Seasonal variability in M2 and M4 tidal constituents and its implications for the coastal residual sediment transport, Geophysical Research Letters, 41, 5563–5570, https://doi.org/https://doi.org/10.1002/2014GL060517, 2014.

Groeneweg, J., Caires, S., van Nieuwkoop, J., and Bottema, M.: A first assessment of the effect of storm climate trends and uncertainties on Dutch levee design, Journal of Flood Risk Management, 15, e12 808, https://doi.org/DOI: 10.1111/jfr3.12808, 2022.

Guo, L., van der Wegen, M., Jay, D. A., Matte, P., Wang, Z. B., Roelvink, D., and He, Q.: River-tide dynamics: Exploration of nonstationary and nonlinear tidal behavior in the Y angtze R iver estuary, Journal of Geophysical Research: Oceans, 120, 3499–3521, 2015.

Guo, L., Wang, Z. B., Townend, I., and He, Q.: Quantification of tidal asymmetry and its nonstationary variations, Journal of Geophysical Research: Oceans, 124, 773–787, https://doi.org/https://doi.org/10.1029/2018JC014372, 2019.

Haigh, I. D., Marcos, M., Talke, S. A., Woodworth, P. L., Hunter, J. R., Hague, B. S., Arns, A., Bradshaw, E., and Thompson, P.: GESLA version 3: A major update to the global higher-frequency sea-level dataset, Geoscience Data Journal, 10, 293–314, https://doi.org/https://doi.org/10.1002/gdj3.174, 2023.

Horsburgh, K. and Wilson, C.: Tide-surge interaction and its role in the distribution of surge residuals in the North Sea, Journal of Geophysical Research: Oceans, 112, https://doi.org/https://doi.org/10.1029/2006JC004033, 2007.

Hsiao, S.-C., Chiang, W.-S., Jang, J.-H., Wu, H.-L., Lu, W.-S., Chen, W.-B., and Wu, Y.-T.: Flood risk influenced by the compound effect of storm surge and rainfall under climate change for low-lying coastal areas, Science of the total environment, 764, 144 439, https://doi.org/https://doi.org/10.1016/j.gloplacha.2005.07.004, 2021.

Katsman, C. A., Sterl, A., Beersma, J., Van den Brink, H., Church, J., Hazeleger, W., Kopp, R., Kroon, D., Kwadijk, J., Lammersen, R., et al.: Exploring high-end scenarios for local sea level rise to develop flood protection strategies for a low-lying delta—the Netherlands as an example, Climatic change, 109, 617–645, https://doi.org/DOI 10.1007/s10584-011-0037-5, 2011.

Kendall, M. G.: A new measure of rank correlation, Biometrika, 30, 81–93, https://doi.org/https://doi.org/10.2307/2332226, 1938.

Kendall, M. G.: Rank correlation methods. Griffin, London, Kendall MG, 1975.

Khodabina, M. and Ahmadabadib, A.: Some properties of generalized gamma distribution, 2010.

Kollu, R., Rayapudi, S. R., Narasimham, S., and Pakkurthi, K. M.: Mixture probability distribution functions to model wind speed distributions, International Journal of energy and environmental engineering, 3, 1–10, https://doi.org/https://doi.org/10.1186/2251-6832-3-27, 2012.

Koutsoyiannis, D., Dimitriadis, P., Lombardo, F., and Stevens, S.: From Fractals to Stochastics: Seeking Theoretical Consistency in Analysis of Geophysical Data, pp. 237–278, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-58895-7_14, 2018.

Lin, N., Emanuel, K. A., Smith, J. A., and Vanmarcke, E.: Risk assessment of hurricane storm surge for New York City, Journal of Geophysical Research: Atmospheres, 115, https://doi.org/10.1029/2009JD013630, 2010.

Liu, J. C., Lence, B. J., and Isaacson, M.: Direct joint probability method for estimating extreme sea levels, Journal of waterway, port, coastal, and ocean engineering, 136, 66–76, https://doi.org/doi.org/10.1061/(ASCE)0733-950X(2010)136:1(66), 2010.

Mann, H. B.: Nonparametric tests against trend, Econometrica: Journal of the econometric society, pp. 245–259, https://doi.org/https://doi.org/10.2307/1907187, 1945.

Medvedev, I., Kulikov, E., and Rabinovich, A.: Tidal oscillations in the Caspian Sea, Oceanology, 57, 360–375, https://doi.org/https://doi.org/10.1134/S0001437017020138, 2017.

Medvedev, I. P., Vilibić, I., and Rabinovich, A. B.: Tidal resonance in the Adriatic Sea: Observational evidence, Journal of Geophysical Research: Oceans, 125, e2020JC016 168, https://doi.org/https://doi.org/10.1029/2020JC016168, 2020.

Munk, W., Zetler, B., and Groves, G.: Tidal cusps, Geophysical Journal International, 10, 211–219, https://doi.org/https://doi.org/10.1111/j.1365-246X.1965.tb03062.x, 1965.

Naderi, A. and Siadatmousavi, S. M.: Extreme value analysis for waves in the Persian Gulf: Skill assessment of different methods for a fetch-limited basin, Regional Studies in Marine Science, 59, 102 812, https://doi.org/https://doi.org/10.1016/j.rsma.2023.102812, 2023.

Neumann, B., Vafeidis, A. T., Zimmermann, J., and Nicholls, R. J.: Future coastal population growth and exposure to sea-level rise and coastal flooding-a global assessment, PloS one, 10, e0118 571, 2015.

Nicholls, R. J.: Coastal megacities and climate change, GeoJournal, 37, 369–379, https://doi.org/https://doi.org/10.1007/BF00814018, 1995.

Nicholls, R. J., Wong, P. P., Burkett, V., Woodroffe, C. D., and Hay, J.: Climate change and coastal vulnerability assessment: scenarios for integrated assessment, Sustainability Science, 3, 89–102, https://doi.org/https://doi.org/10.1007/s11625-008-0050-4, 2008.

Nienhuis, J. H. and Van de Wal, R. S.: Projections of global delta land loss from sea-level rise in the 21st century, Geophysical Research Letters, 48, e2021GL093 368, https://doi.org/https://doi.org/10.1029/2021GL093368, 2021.

Paalvast, P.: Ecological studies in a man-made estuarine environment, the port of Rotterdam, Ph.D. thesis, [Sl]:[Sn], https://doi.org/10.13140/2.1.1101.6486, 2014.

Papalexiou, S., Koutsoyiannis, D., and Makropoulos, C.: How extreme is extreme? An assessment of daily rainfall distribution tails, Hydrology and Earth System Sciences, 17, 851–862, https://doi.org/https://doi.org/10.5194/hess-17-851-2013, 2013.

Pappas, C., Papalexiou, S. M., and Koutsoyiannis, D.: A quick gap filling of missing hydrometeorological data, Journal of Geophysical Research: Atmospheres, 119, 9290–9300, https://doi.org/https://doi.org/10.1002/2014JD021633, 2014.

Pattiaratchi, C.: Coastal tide gauge observations: dynamic processes present in the Fremantle record, Operational oceanography in the 21st century, pp. 185–202, https://doi.org/https://doi.org/10.1007/978-94-007-0332-2_7, 2011.

Pawlowicz, R., Beardsley, B., and Lentz, S.: Classical tidal harmonic analysis including error estimates in MATLAB using T_TIDE, Computers & Geosciences, 28, 929–937, https://doi.org/https://doi.org/10.1016/S0098-3004(02)00013-4, 2002.

Prestes, Y. O., Silva, A. C., Rollnic, M., and ROSÁRIO, R. P.: The M2 and M4 tides in the Pará River estuary, Tropical Oceanography, 45, 1679–3013, https://doi.org/10.5914/tropocean.v45i1.15198, 2017.

Proudman, J.: The effect of friction on a progressive wave of tide and surge in an estuary, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 233, 407–418, https://doi.org/https://www.jstor.org/stable/99898, 1955a.

Proudman, J.: The propagation of tide and surge in an estuary, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 231, 8–24, https://doi.org/https://doi.org/10.1098/rspa.1955.0153, 1955b.

Pugh, D. T. and Vassie, J.: Extreme sea levels from tide and surge probability, in: Coastal Engineering 1978, pp. 911–930, https://doi.org/https://doi.org/10.9753/icce.v16.52, 1978.

Ragno, E., Antonini, A., and Pasquali, D.: Investigating extreme sea level components and their interactions in the Adriatic and Tyrrhenian Seas, Weather and Climate Extremes, 41, 100 590, https://doi.org/https://doi.org/10.1016/j.wace.2023.100590, 2023.

Rossiter, J. R.: Interaction between tide and surge in the Thames, Geophysical Journal International, 6, 29–53, https://doi.org/https://doi.org/10.1111/j.1365-246X.1961.tb02960.x, 1961.

Sánchez-Arcilla, A., Jiménez, J. A., Valdemoro, H. I., and Gracia, V.: Implications of climatic change on Spanish Mediterranean low-lying coasts: The Ebro delta case, Journal of Coastal Research, 24, 306–316, https://doi.org/https://doi.org/10.2112/07A-0005.1, 2008.

Scarrott, C. and MacDonald, A.: A review of extreme value threshold estimation and uncertainty quantification, REVSTAT-Statistical journal, 10, 33–60, https://doi.org/https://doi.org/10.57805/revstat.v10i1.110, 2012.

Schmitt, F. G., Crapoulet, A., Hequette, A., and Huang, Y.: Nonlinear dynamics of the sea level time series in the eastern English Channel, Natural Hazards, 91, 267–285, https://doi.org/https://doi.org/10.1007/s11069-017-3125-7, 2018.

Simon, B.: Coastal tides, Institut océanographique éd., 2013.

Sklar, A.: Random variables, joint distribution functions, and copulas, Kybernetika, 9, 449–460, https://doi.org/http://dml.cz/dmlcz/125838, 1973.

Slomp, R., Knoeff, H., Bizzarri, A., Bottema, M., and de Vries, W.: Probabilistic flood defence assessment tools, in: E3S Web of conferences, vol. 7, p. 03015, EDP Sciences, https://doi.org/10.1051/e3sconf/20160703015, 2016.

Small, C. and Nicholls, R. J.: A global analysis of human settlement in coastal zones, Journal of coastal research, pp. 584–599, https://doi.org/https://www.jstor.org/stable/4299200, 2003.

Spicer, P., Huguenard, K., Ross, L., and Rickard, L. N.: High-frequency tide-surge-river interaction in estuaries: Causes and implications for coastal flooding, Journal of Geophysical Research: Oceans, 124, 9517–9530, https://doi.org/https://doi.org/10.1029/2019JC015466, 2019.

Stijnen, J., Kanning, W., Jonkman, S., and Kok, M.: The technical and financial sustainability of the Dutch polder approach, Journal of Flood Risk Management, 7, 3–15, https://doi.org/https://doi.org/10.1111/jfr3.12022, 2014.

Teng, F., Fang, G., Xu, X., and Zhu, Y.: Shallow water tides induced by frictional nonlinearity in the Bohai and Yellow Seas, Journal of Sea Research, 194, 102 411, https://doi.org/https://doi.org/10.1016/j.seares.2023.102411, 2023.

Thomson, R. E. and Emery, W. J.: Data analysis methods in physical oceanography, Newnes, 2014.

Van Alphen, J., Haasnoot, M., and Diermanse, F.: Uncertain accelerated sea-level rise, potential consequences, and adaptive strategies in the Netherlands, Water, 14, 1527, https://doi.org/https://doi.org/10.3390/w14101527, 2022.

Van den Brink, H.: Extreme wind en druk in de ECMWF seizoensverwachtingen, Report, KNMI, de Bilt (in Dutch), 2018.

Van den Brink, H. W. and de Goederen, S.: Recurrence intervals for the closure of the Dutch Maeslant surge barrier, Ocean Science, 13, 691–701, https://doi.org/10.5194/os-13-691-2017, 2017.

Wahl, T., Haigh, I. D., Nicholls, R. J., Arns, A., Dangendorf, S., Hinkel, J., and Slangen, A. B.: Understanding extreme sea levels for broad-scale coastal impact and adaptation analysis, Nature communications, 8, 16 075, https://doi.org/https://doi.org/10.1038/ncomms16075, 2017.

Weisse, R., von Storch, H., Niemeyer, H. D., and Knaack, H.: Changing North Sea storm surge climate: An increasing hazard?, Ocean & coastal management, 68, 58–68, https://doi.org/https://doi.org/10.1016/j.ocecoaman.2011.09.005, 2012.

Williams, J., Horsburgh, K. J., Williams, J. A., and Proctor, R. N.: Tide and skew surge independence: New insights for flood risk, Geophysical Research Letters, 43, 6410–6417, https://doi.org/10.1002/2016GL069522., 2016.

Woodruff, J. D., Irish, J. L., and Camargo, S. J.: Coastal flooding by tropical cyclones and sea-level rise, Nature, 504, 44–52, https://doi.org/https://doi.org/10.1038/nature12855, 2013.

Woodworth, P. L., Melet, A., Marcos, M., Ray, R. D., Wöppelmann, G., Sasaki, Y. N., Cirano, M., Hibbert, A., Huthnance, J. M., Monserrat, S., et al.: Forcing factors affecting sea level changes at the coast, Surveys in Geophysics, 40, 1351–1397, https://doi.org/https://doi.org/10.1007/s10712-019-09531-1, 2019.

Yue, S. and Wang, C.: The Mann-Kendall test modified by effective sample size to detect trend in serially correlated hydrological series, Water resources management, 18, 201–218, https://doi.org/https://doi.org/10.1023/B:WARM.0000043140.61082.60, 2004.

Yue, S. and Wang, C. Y.: Applicability of prewhitening to eliminate the influence of serial correlation on the Mann-Kendall test, Water resources research, 38, 4–1, https://doi.org/https://doi.org/10.1029/2001WR000861, 2002.

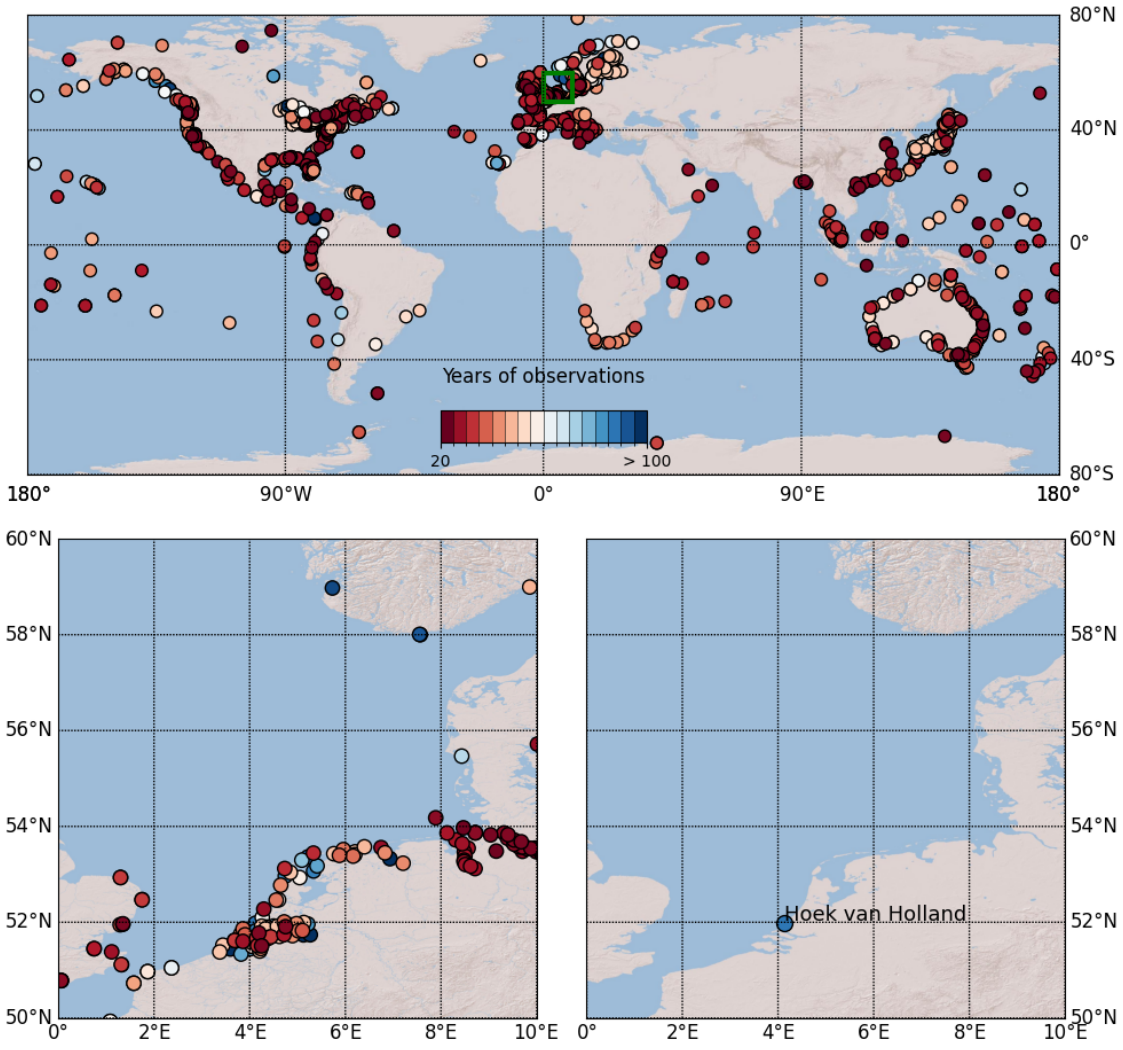
II

Supplement

Supplement A: Data

A1 GESLA spatial distribution

Global Extreme Sea Level Analysis (GESLA) contain water level data for 5119 stations worldwide. In our study case the data of Hook of Holland station are taken into consideration. The worldwide spatial distribution of GESLA data stations with more than 20 years of observations can be found in Sl Figure A1, as well as the specific coordinates of the station of interest, latitude: 51°98.0'N, longitude: 4°12.0'E.



Sl Figure. A1. Locations of all water level gauges in the Global Extreme Sea Level Analysis (GESLA) dataset.

A2 Comparison between GESLA and RWS data

We investigated if the GESLA data of 66 years of observations (1953-2018) can be representative to the RWS longer data (1887-2020) of 134 years, based on results of EVA for the threshold of $212.2cm$ and declustering time of 4 days. For that reason, from RWS data, 1000 subsamples with 66 years of observations are generated and the boxplots of their goodness of fit scores are included in Sl Figure A2. Additionally, the water levels with a probability of occurrence equal to $1/10000years$ are given. From the Kolmogorov-Smirnov test (K-S) graph, it is observed that all the distributions lead to small errors. Apart of this, the lowest errors are achieved by Truncated Gumbel and Burr XII distributions for both RWS and GESLA data, while for RWS data the error of Pareto distribution is also quite small. In all the distributions the value of K-S from the whole RWS dataset is between the $25th$ and $75th$ percentiles of boxplot while the GESLA are between $5th$ and $95th$. For Root Mean Square Error (RMSE), again all the distributions have quite small errors in the estimation of Extreme Water Level (EWL). We see that the smallest boxes correspond to Gumbel, Generalized Pareto, and Burr XII distributions, which means that they produce small errors in any of the random subsamples. The smaller errors are given for RWS data, by Burr XII and Generalized Pareto, but the difference between them and Gumbel and Generalized Gamma are almost negligible. According to the estimation of EWL with a Return Period (RP) of $10000years$ the results from RWS and GESLA data are similar in all the distributions except for Weibull. The latter one overestimates the water level as in RWS data is the only distribution that leads to an estimation of $EWL > 4.5m$ and more specifically $EWL \approx 4.6m$ while is estimated from Gamma $EWL \approx 4.48m$. The overestimation is even larger in GESLA data leading to $EWL \approx 5.4m$ the only estimation above $5m$. One possible explanation for this variation and overestimation could be that is part of "light tail" distributions and very sensitive to the number of samples, especially in GESLA data. On the other hand, the estimation of Gumbel and Burr XII distributions that had the best scores in the goodness of fit parameters, are almost identical with $EWL_{Gumbel,RWS} = 4.43m$, $EWL_{Burr,RWS} = 4.40m$ for RWS and $EWL_{Gumbel,GESLA} = 4.36m$, $EWL_{Burr,GESLA} = 4.32m$, respectively.

Summarizing all the above results, the GESLA dataset can successfully represent the longer RWS data.

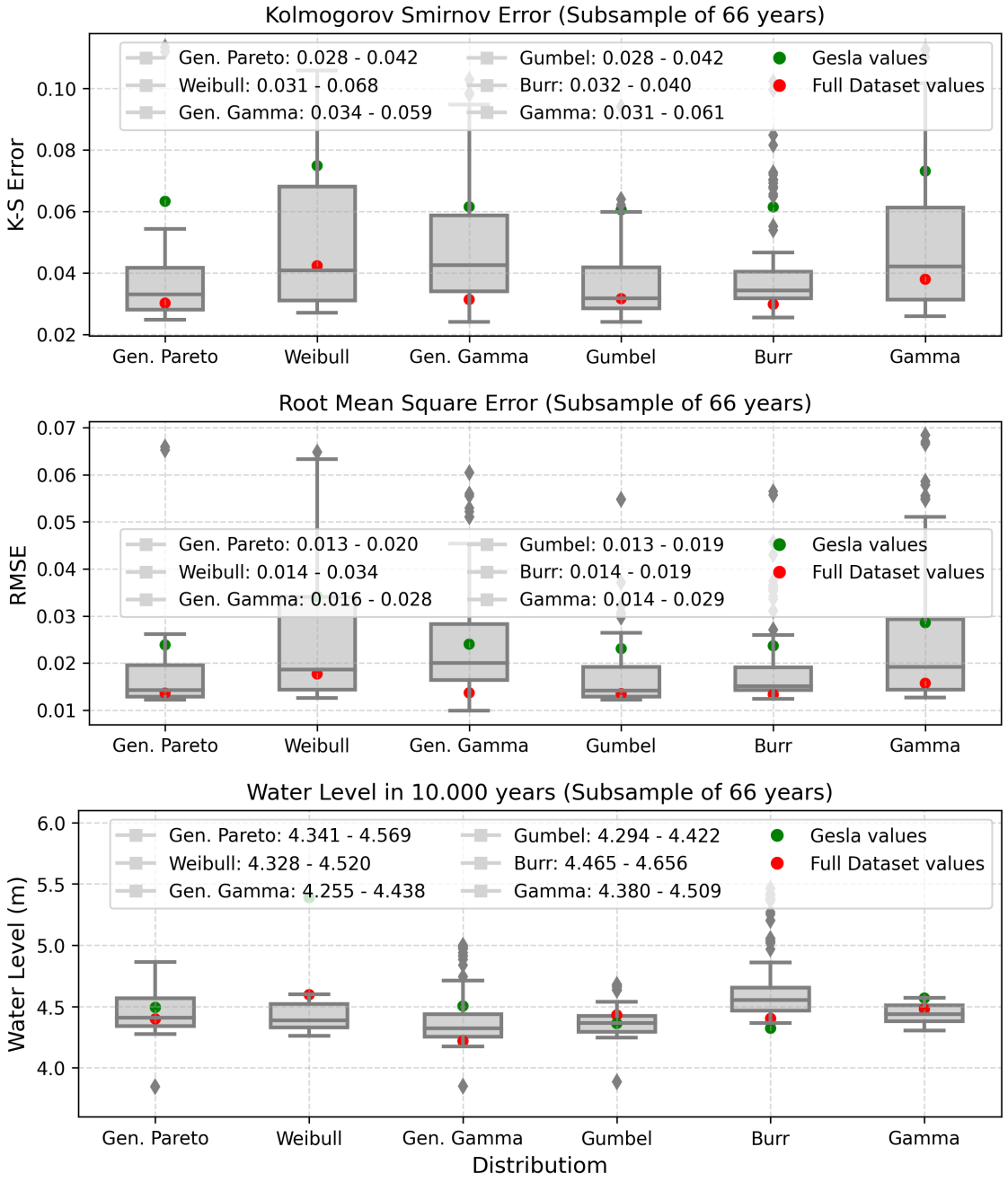
Supplement B: Mann Kendall Test Trend Analysis

B1 SLR effect in the intensity of Extreme Water Levels

Trend analysis, apart from the homogenization of data by removing their linear trends and correcting them to the final year of observations, can also focus on the effect of SLR on the intensity of Extreme Water Levels. For that reason, the longer RWS data are used. The annual maxima and annual mean sea level of the measurements were identified and the 19-year moving average have been calculated for trend detection without the effect of the lunar nodal cycle (18.6 years). Studying the trend of the values (Sl Figure.B1) from a visual inspection of the 19 years an increasing trend can be seen. Using Mann-Kendall and pre-whitening Mann-Kendall Tests this visual observation is validated as for the yearly maxima both tests reject the hypothesis of the absence of a trend at the 95% confidence level ($pvalue = 0.006$ for Mann-Kendall and $pvalue = 0.02$ for the pre-whitening test, respectively). For determination of the influence of mean sea level rise in this result, the yearly mean sea level is subtracted from the yearly maxima to verify if any trend continues to exist. From the results of the tests ($pvalue = 0.11$ for both tests) the null hypothesis of no trend cannot be rejected, supporting the theory that sea level rise is the main factor for the increase in annual water level.

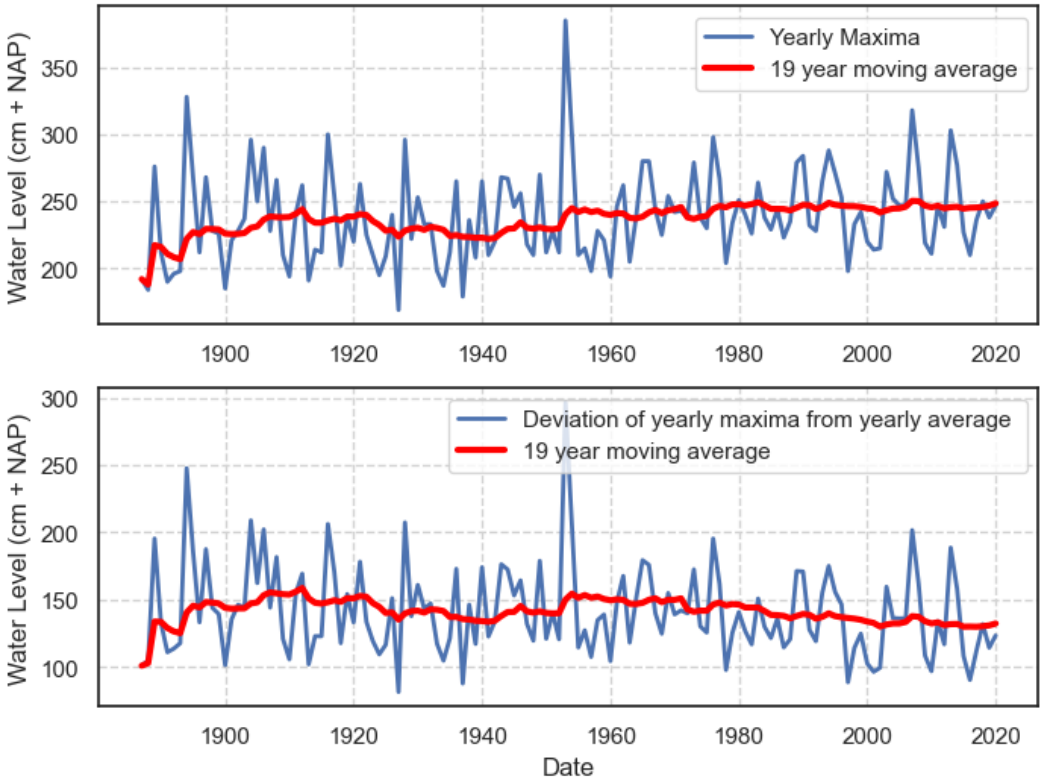
B2 Detrending of RWS data

To remove any trend, and jumps of RWS data to make them homogeneous. The yearly mean water levels of RWS data (Sl Figure. B2) show jumps connected with man-made interventions and increasing trends from the rising of MSL. The data present no trend ($pvalue = 1.0$ for both tests) until the first breakpoint in 1989 which is contrary to the increasing one from Caires (2011), probably because in our dataset only the two highest water level per day are included. After that, an increasing



Sl Figure. A2. Kolmogorov-Smirnov test (K-S), Root Mean Square Error (RMSE) and estimation of extreme water level with a probability of occurrence $1/10000 years$ for the distributions that have been chosen for Extreme Value Analysis.

trend from 1890 to 1900 ($pvalue = 0.029$ for both tests) is detected followed by another increasing trend ($pvalue = 0.006$ for Mann-Kendall and $pvalue = 0.04$ for pre-whitening test, respectively) with milder slope until 1965 when a jump proba-



Sl Figure. B1. The yearly maxima and the yearly maxima subtracting the yearly mean sea level are presented for the study of their trends. From visual inspection of the 19-year moving average, a clearly increasing trend is shown for annual maximum values and an absence trend for the second graph, observations confirmed by the Mann-Kendall Test results.

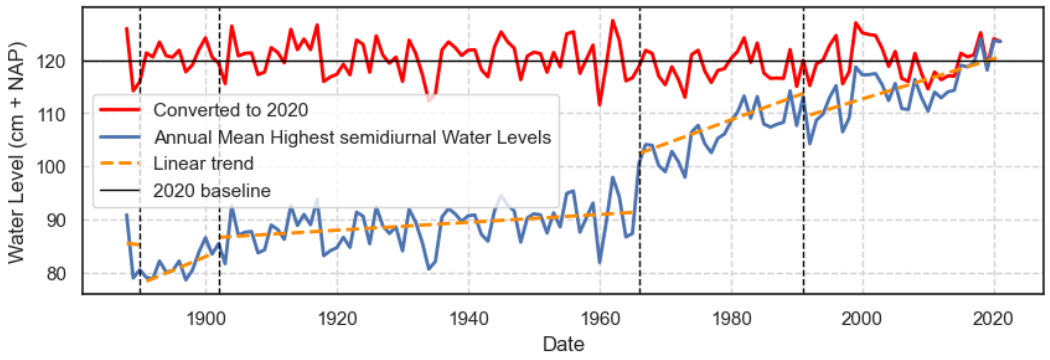
bly connected with the works for Maasvlakte1 and industrial area (Paalvast, 2014; Caires, 2011) is observed. For the period 1965-1990, the trend is strongly increasing ($pvalue = 0.00006$ for Mann-Kendall and $pvalue = 0.033$ for the pre-whitening test, respectively), followed by a sharp decline in 1990 caused by the connection Harteelkanaal- Beerkanaal. From the end of 1990 to 2020 the trend is still increasing ($pvalue = 0.0002$ for Mann-Kendall and $pvalue = 0.046$ for the pre-whitening test, respectively), but with a milder slope than previously, which could mean that the sea level rise effect is less evident.

Supplement C: Sea Surface Pressure and Wind speed

In this chapter, the spectrum of sea surface pressure is presented, and an analysis of the correlation between extreme Non Tidal Residuals (NTR) and wind speed is implemented to investigate the influence of threshold in correlation heatmap. Additionally, the direction of maximum wind speeds correlated with the extreme NTR is shown.

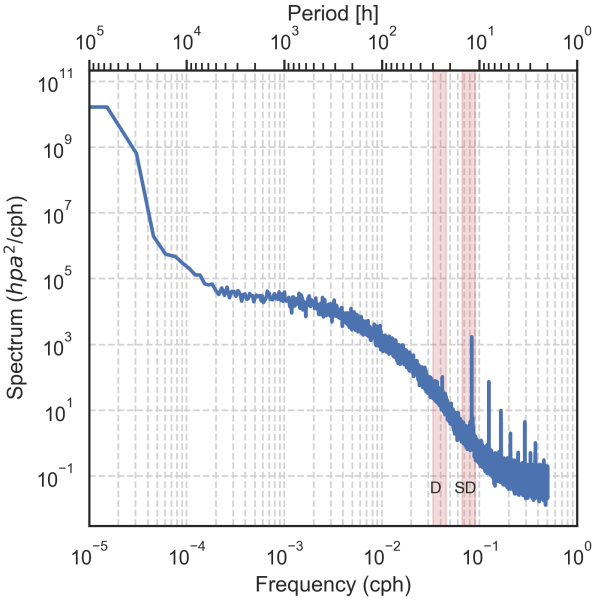
C1 Sea surface pressure spectral analysis

The spectrum of surface pressure has been analyzed presenting a predominance of semi-diurnal surface pressure variation, similar to spectral analysis of water level (Sl Figure. C1). Moreover, the second highest peak that is reached in a frequency of



Sl Figure. B2. Illustration of the annual mean of highest semidiurnal water levels from RWS data. The linear trends have been highlighted by the dashed orange line. The annual mean of homogenized and corrected data to the last year of observations (2020) is represented by the red line.

$freq = 0.125cph$, or in a period of $T = 8h$ can be responsible for the high coherence between sea surface pressure and NTR in this frequency region.



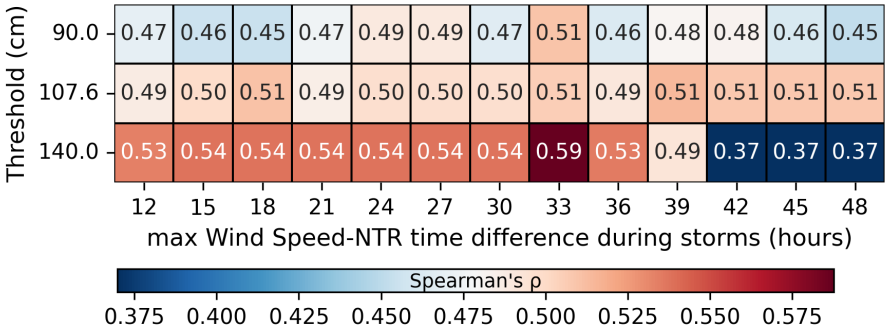
Sl Figure. C1. Spectral Analysis of Surface Pressure Data. The red shaded areas determine the diurnal (D) and semidiurnal (SD) areas.

C2 Wind speed analysis

C2.1 Wind speed analysis-NTR correlation

For the investigation of the influence of the threshold of extreme NTR on both the value of the correlation, as well as the time that this happens for a delcustering time of $4days$, different values of threshold are evaluated to compare their results. All the values of NTR above the threshold correspond with the critical direction of $340°$. The selection is made with percentile criteria so the three selected thresholds are, the $99th$ percentile $90cm$, the $99.5th$ percentile $107,6cm$, and the $140cm$ as a reference point. From the results, we see that a higher threshold leads to a higher correlation (Sl Figure C2).

For the smallest threshold, the maximum correlation $\rho = 0.51$ is reached in the window of $33hours$, while the threshold of $107,6cm$ gives the highest correlation $\rho = 0.51$ in $18hours$, but the same correlation is also reached in $33, 39, 42, 45and48hours$. The maximum correlation for the threshold of $140cmcm$ is the highest $\rho = 0.59$ and is reached at $33hours$.



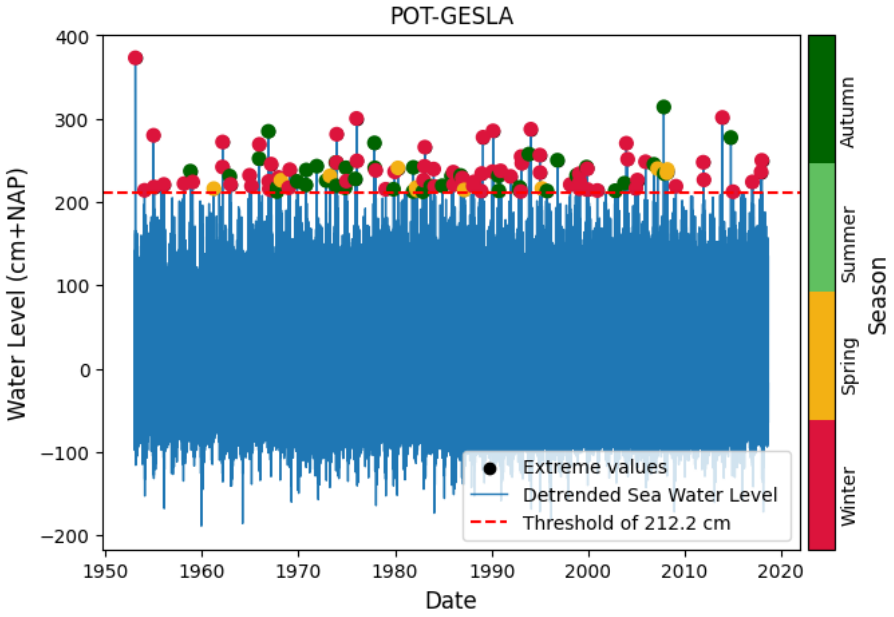
Sl Figure. C2. Heatmap of the correlation between NTR and wind speed peaks for the critical wind direction of $340°$ for different threshold values of NTR.

Supplement D: Extremes

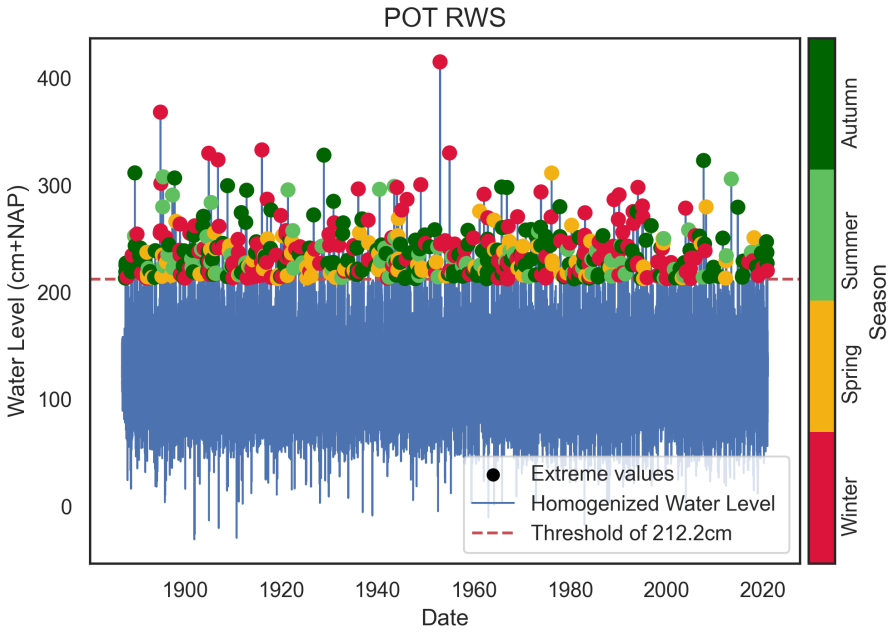
D1 Peak Over Threshold (POT)

The peak values of EWL above the selected threshold of $212.2cm$ and the declustering time of $4days$ are given in Sl Figure. D1 for GESLA and Sl Figure. D2 for RWS data respectively. From the seasonal distribution of both of the datasets can be clearly seen that the majority of extreme observations is captured in winter period. According to the number of extreme events per dataset, for the GESLA data 124 events have been extracted after the POT method ($1.88event/year$). On the other hand for RWS data, 525 events have been found ($3.83event/year$). Two possible explanations can be given for the above result 1) the number of extreme events was higher in the period 1887-1952 than later, so this phenomenon increases the total number of peaks. Or 2) There is a difference in the measurements of GESLA and RWS data and this is leading to systematically higher water levels of RWS compared to GESLA data. To distinguish the real reason for the variation in peak number we calculated the number of peaks for RWS data for the period 1953-2018 which is the same as the period of GESLA data. $241peaks$ are detected or ($3.68event/year$) which means that there is a possible difference in the measurements between GESLA and RWS data that makes the latter to give higher EWL in the same events.

Sensitivity analysis is important to be implemented to understand how the number of peaks above threshold changes, based on different thresholds and declustering times. Results of this analysis can be seen in Sl Figure. D3, D4, for GESLA and in Sl



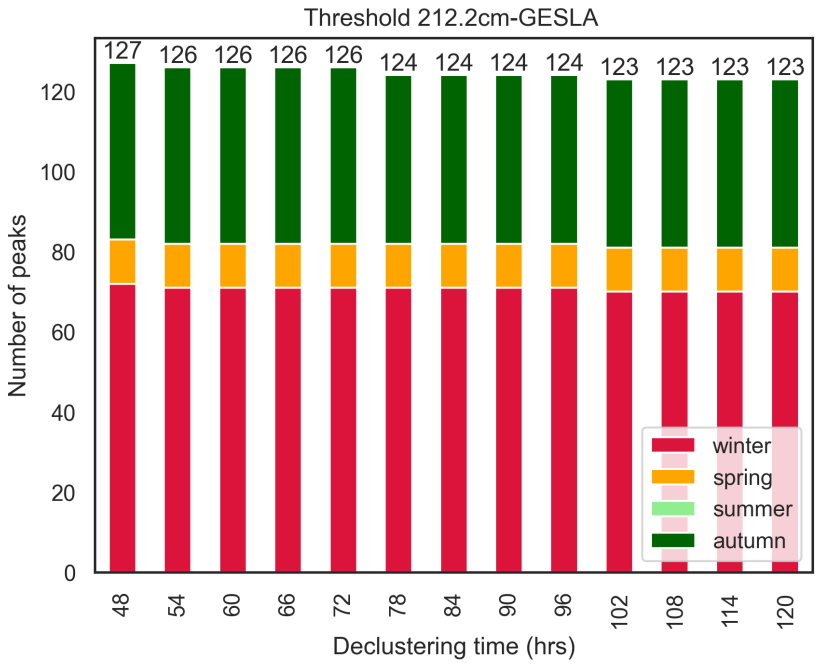
Sl Figure. D1. Seasonal distribution of the peaks after POT method for GESLA data .



Sl Figure. D2. Seasonal distribution of the peaks after POT method for RWS data .

Figure. D5, D6, for RWS data. We observe that the declustering time doesn't affect so much the number of peaks above the

threshold. On the other hand, the impact of the threshold is decisive for the number of extreme peak events. So the POT method for our data is much more sensitive to the changes in the threshold, than the declustering time. As can be seen from Sl Figure. D4, D6 as the initial threshold is small $(140cm)$ all the seasons produce almost the same number of peaks but as the threshold increases the dominance of the winter season is obvious, as the most extreme surges happen in that season. Additionally, the second most influenced period for EWL is the autumn season, with only a few events to correspond with the spring season.

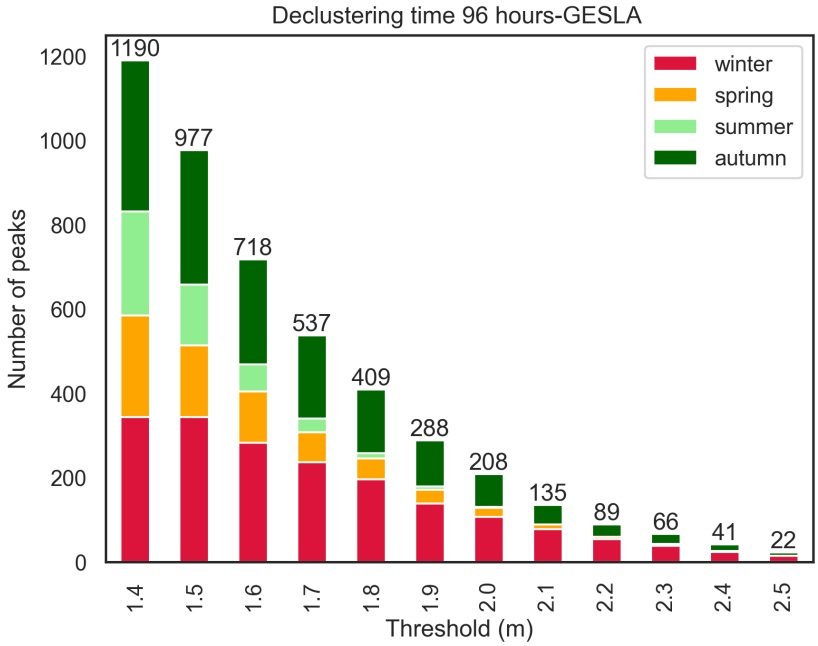


Sl Figure. D3. Number of peaks above the threshold for GESLA data based on different declustering times, from 2 to 5 days. For a threshold of 212.2cm

D2 Extreme Value Analysis (EVA)

In this subchapter, the main results of EVA for both GESLA and RWS data are presented. In the Sl Figure. D7, D8 the CDF and PDF fitting for GESLA data is illustrated, while in Sl Figure. D9, D10 the same results for RWS data are shown. It is clearly seen that for RWS data the variability of EWL, between the different distributions is much smaller compared to the GESLA data that are shorter, and the less number of data probably influences, especially the "light tail" distributions leading to high variability of the estimations.

Apart from the fitting of distributions, also sensitivity analysis to investigate the influence of threshold selection in the values of the goodness of fit parameters and the estimation of EWL with probability of occurrence $1/10000years$ has been implemented. The declustering time remains at 4 days and the RWS have been used as they are longer and with a larger number of extreme water levels. From the results in Sl Figure. D11, we see that as the threshold increases the scores of Burr XII and Generalized Gamma distribution in goodness of fit parameters are much better than the other 4 distributions. In the case of Burr XII this was expected as it is focused on extreme values but the Generalized Gamma is a "light tail" distribution. Bearing in mind the classes of the distributions it could be assumed that the one extra parameter that they have compared to the other distributions makes them more stable in higher values of threshold. On the other hand, this extra parameter makes them



Sl Figure. D4. Number of peaks above the threshold for GESLA data based on different values of the threshold for a declustering time of 4 days.

more complicated and not as easy to use as the other 4. Observing the estimation of EWL with a probability of occurrence $1/10000 years$ we see some peaks in the estimation for the different thresholds in both Burr XII and Generalized Gamma distributions which means that they are extremely sensitive to the changes of a threshold above the threshold value of $\approx 217cm$. This means that even based on the K-S and RMSE scores it seems that prevail over the others, CDF and PDF distribution fitting graphs are needed to validate these scores by visual inspection.

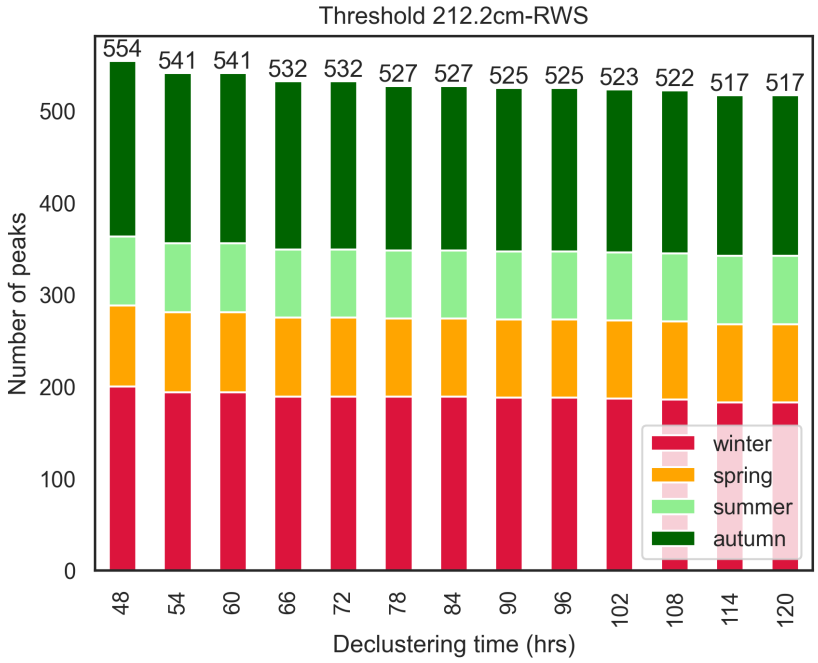
D3 JPM Method

The selection of the CDFs of the two distributions for convolution method is based on their fitting on the extreme NTR above the threshold of $140cm$. As can be seen from the graph the estimations of Burr XII vary compared to all the others so it is chosen, as well as the Generalized Pareto distribution as a representative of the other 4 distributions (Sl Figure. D13). According to the empirical distribution of tides, the Kernel Density Estimation (KDE) is used. As it is seen from the graph the majority of tides are negative (Sl Figure. D14).

D4 Copulas

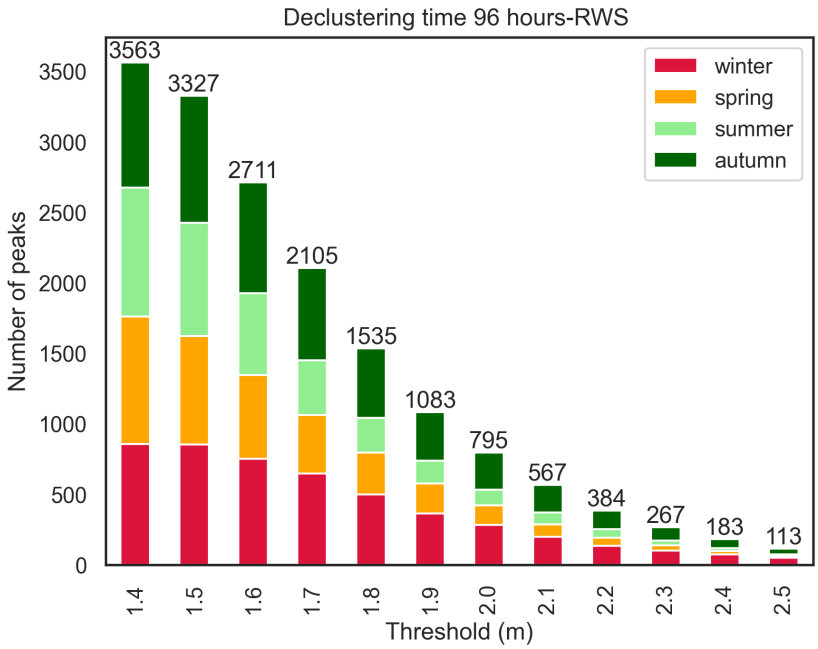
For the copulas method, the fitting of given distributions to the NTR and tides can be found in Sl Figure. D15, D16. As can be seen from the graphs the fitting between the values and the distributions is poor so they cannot be assumed as representative and as a result the empirical distributions were proffered for the copulas method, as marginal distributions.

For the comparison of dependent and independent copulas, as well as the convolution method, a graph is generated (Sl Figure. D17). We can see that the independence copula and the Gaussian one are starting to diverge at low quantiles, while the

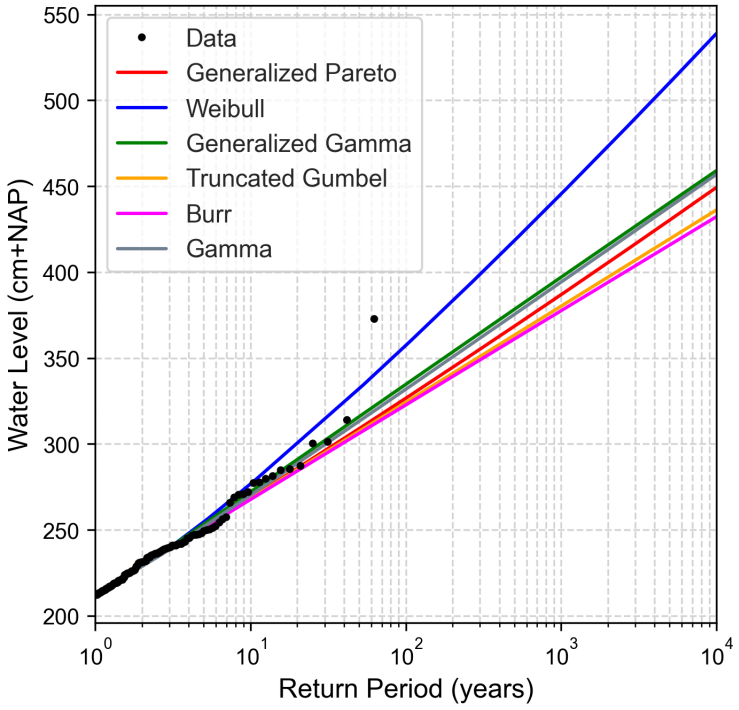


Sl Figure. D5. Number of peaks above the threshold for RWS data based on different declustering times, from 2 to 5 days. For a threshold of 212.2cm

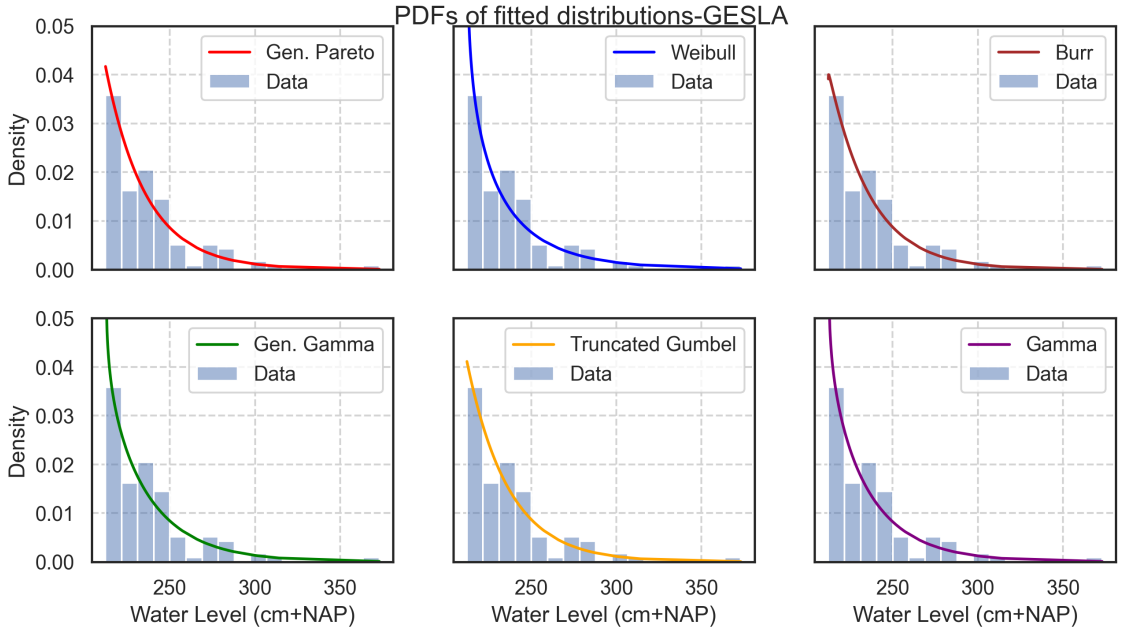
student copula (with the best AIC score) fits better on the observations. On the other hand, we see the poor fitting of convolution distributions that fit to the data only for values of water level above $260cm$ (extremely high quantiles) while for smaller values they diverge from real observations.



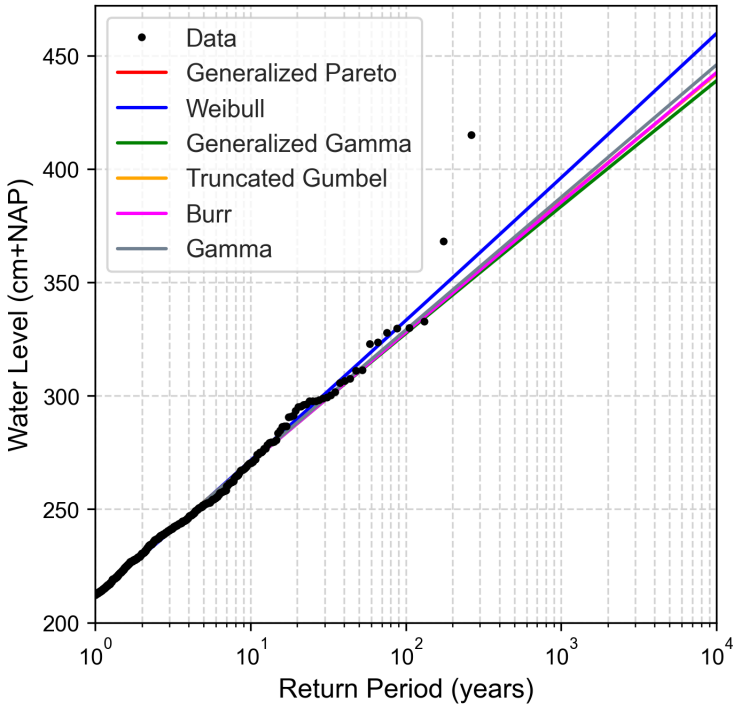
Sl Figure. D6. Number of peaks above the threshold for RWS data based on different values of the threshold for a declustering time of 4 days.



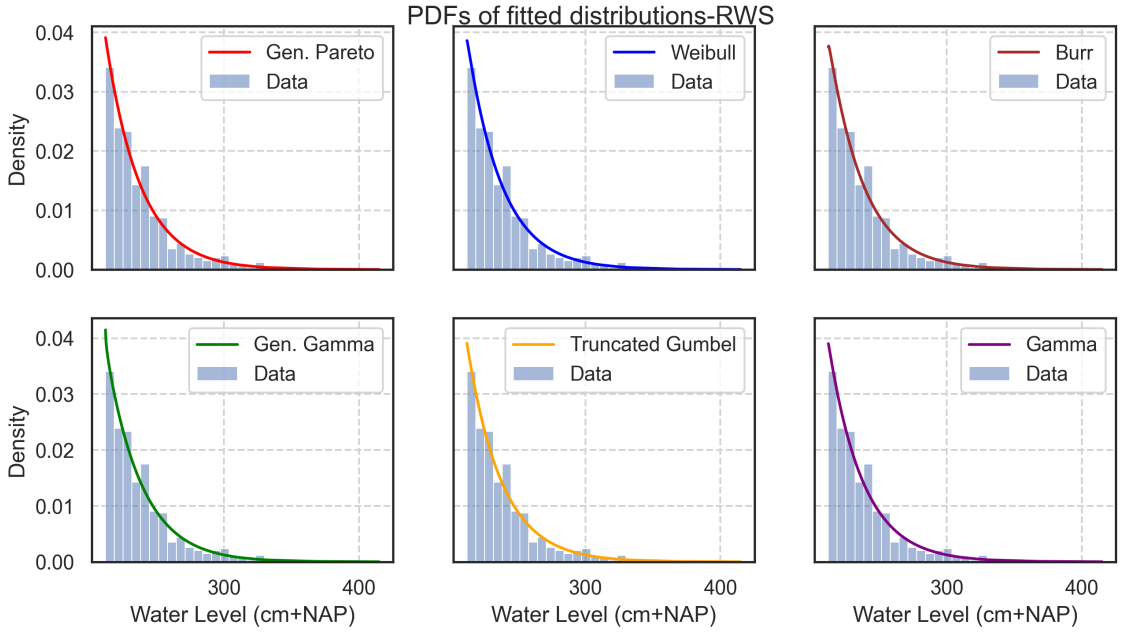
Sl Figure. D7. CDF fitting of given distributions for GESLA data



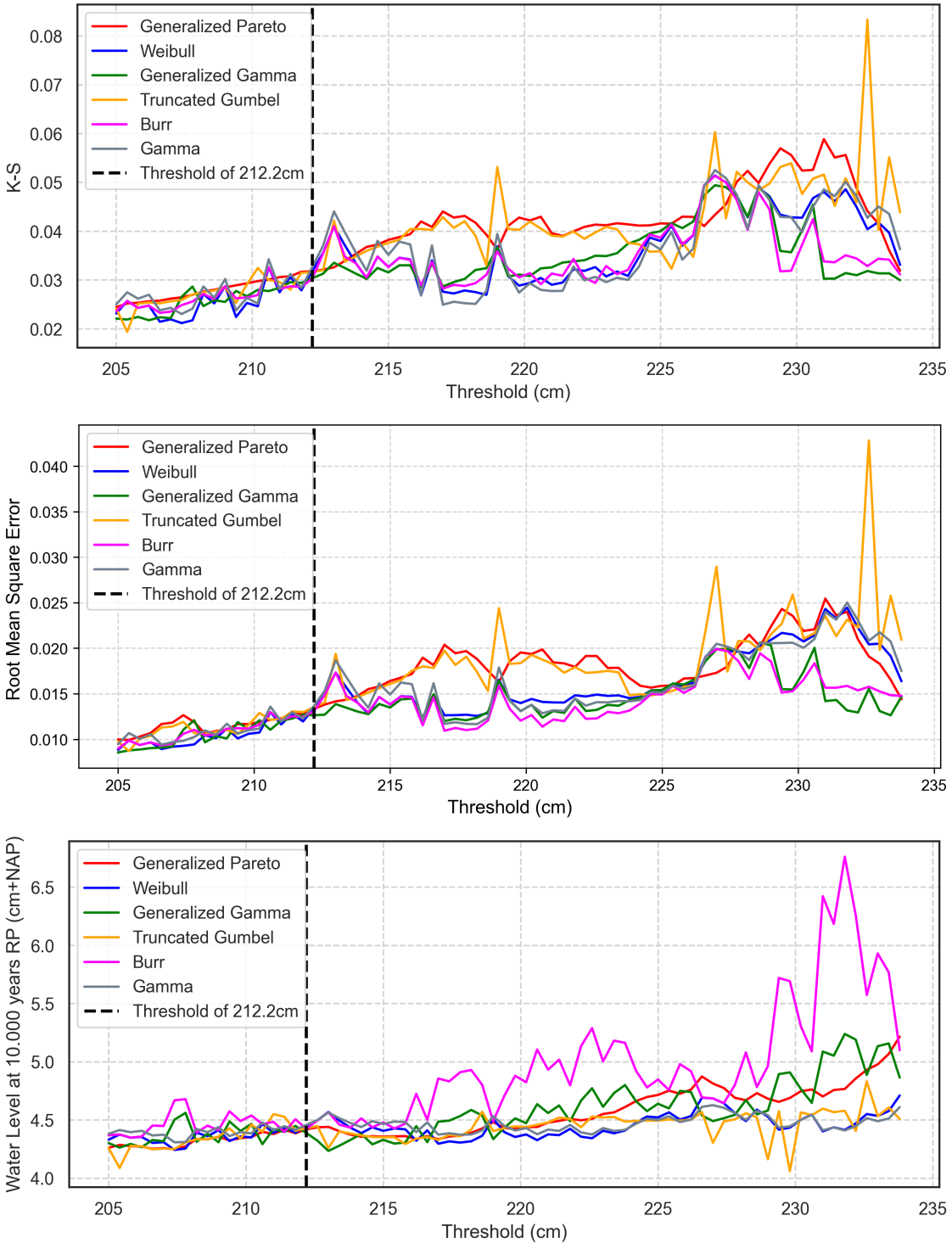
Sl Figure. D8. PDF fitting of given distributions for GESLA data



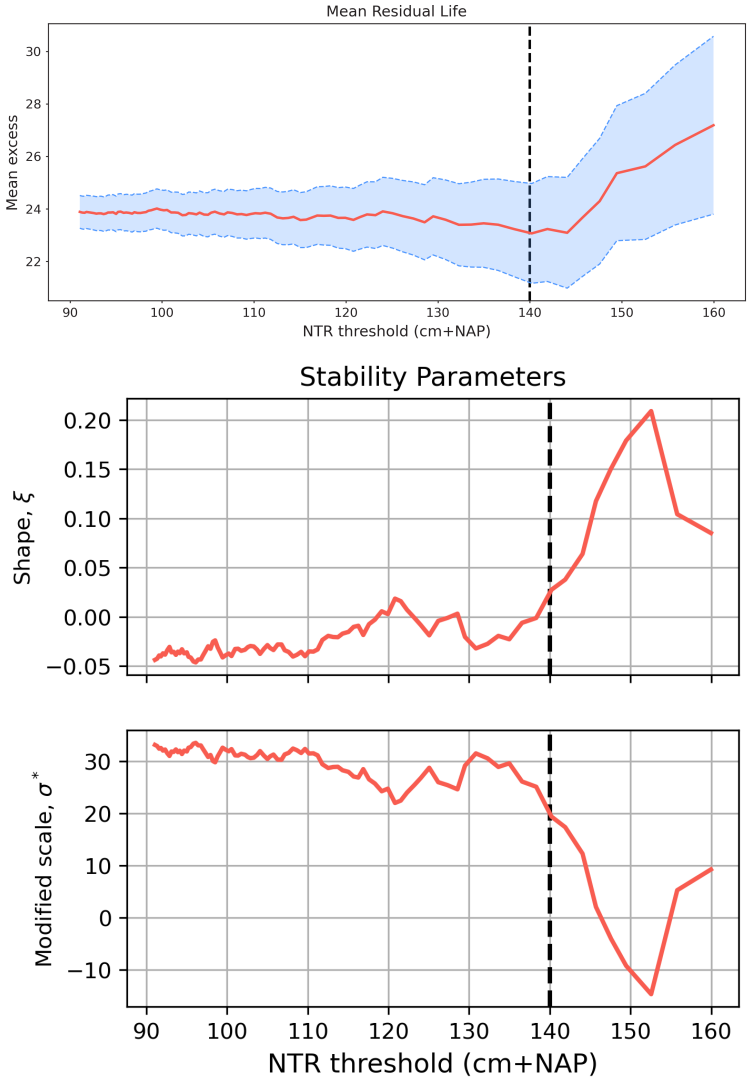
Sl Figure. D9. CDF fitting of given distributions for RWS data



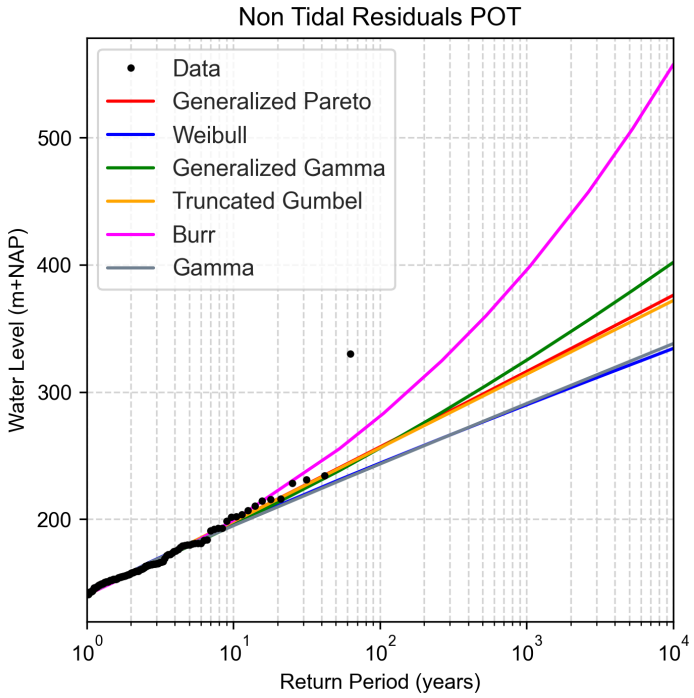
Sl Figure. D10. PDF fitting of given distributions for RWS data



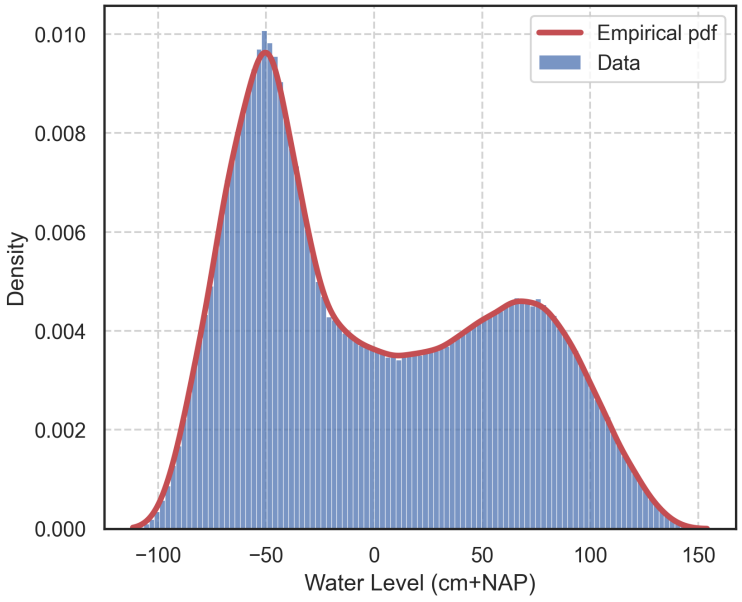
Sl Figure. D11. Sensitivity analysis for the influence of threshold selection in K-S, RMSE parameters and in the estimation of water level with a Return Period of $10000 years$, for the RWS data.



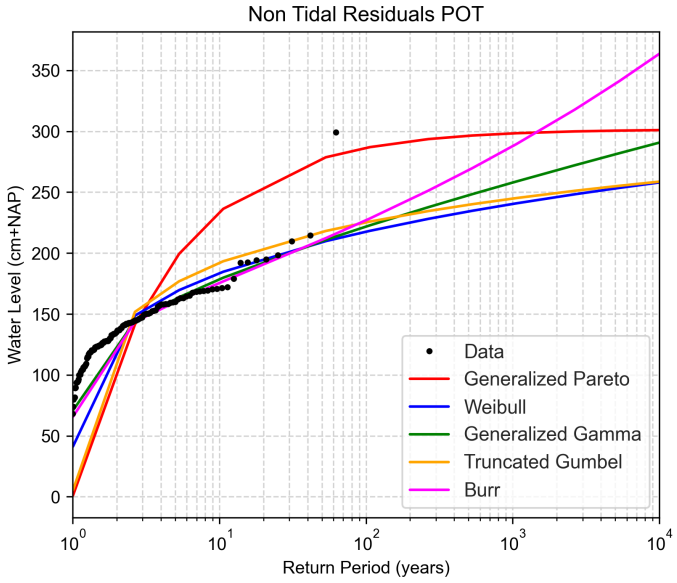
Sl Figure. D12. Mean residual life plot stability plots for NTR in convolution method.



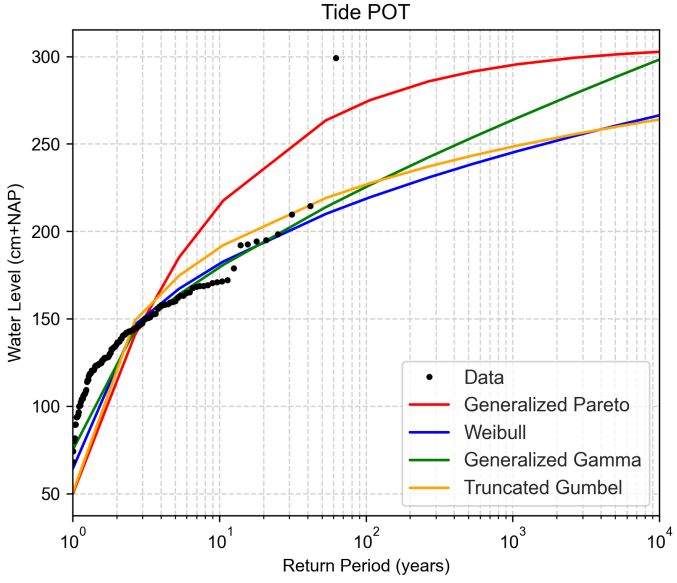
Sl Figure. D13. Fitting of CDF distributions in extreme NTR for convolution method.



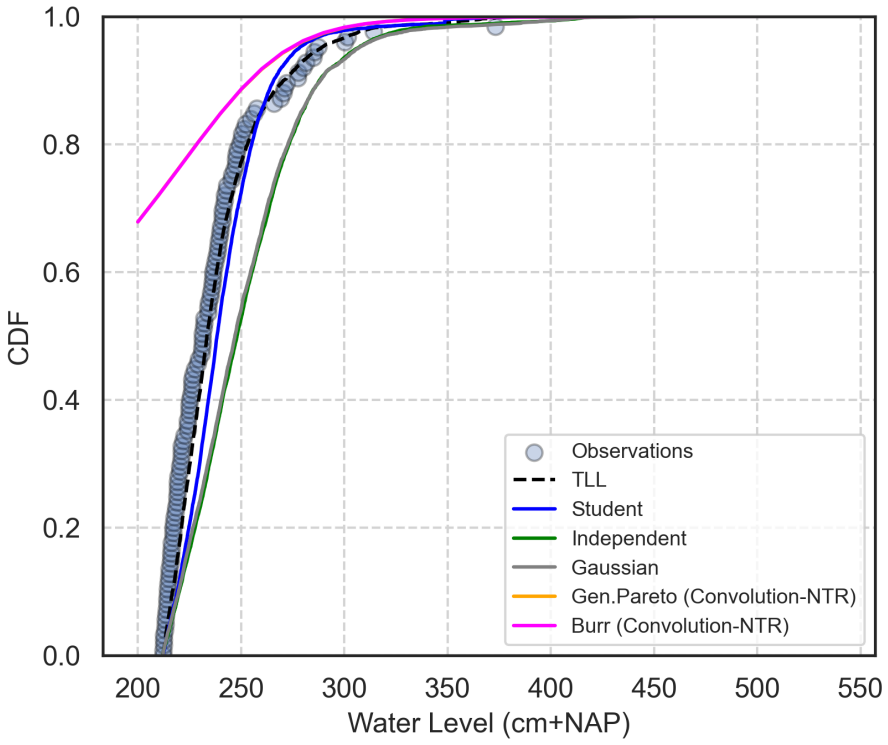
Sl Figure. D14. Empirical PDF of tides for convolution method.



Sl Figure. D15. Fitting of CDF distributions in extreme NTR for copulas.



Sl Figure. D16. Fitting of CDF distributions in extreme tides for copulas.



Sl Figure. D17. Fitting of CDF distributions in extreme tides for copulas.