*An Extreme Analysis and Temporal Clustering of Sea Level Surge Exceedances Around the Irish Coastline.*

**1. Introduction**

**1.1 Background**

**[Provide a brief overview of the topic or problem being addressed in the report. Explain its relevance, significance, and any relevant contextual information.]**

The Irish Coastline stretches approximately 7400 kilometres long and it is estimated that 40% of the Irish population live within 5 kilometres of the coastline (Climate Ireland, CSO Census 2016). This represents a massive risk to communities in coastal areas. Heavy flooding is affecting Irish communities on a near-annual basis. Coastal erosion threatens shorelines, habitats and ecosystems(ICPSS, 2020). These risk factors contribute to infrastructural, agricultural and ecological damage which likely averages to greater than €50 million per year in economic costs (Environmental Protection Agency). Surprisingly, relatively limited analysis has been done to address the issue of extreme sea level surges, despite lots of work being done on flooding more generally. There are some possible reasons for this. There could be a perception of lower risk by comparison to other countries leading to underinvestment in the area. Ireland’s population centres can be prone to riverine flooding and heavy rainfall, potentially diverting attention from coastal surges. The tidal gauge coverage around Ireland is quite good, however a short history of data over time also indicates difficulties is collecting thorough data. There may also be institutional and policy gaps at play, as outlined in a 2023 report by the *Department of Housing, Local Government and Heritage*, many of Ireland’s coasts are fragmented in management by local authorities, and policy is lagging behind the growing understanding of climate impacts, leading to insufficient integration of storm surge analysis in planning and development guidelines

Storm Ophelia in 2017 is considered to be the worst storm in Ireland in 50 years, and the easternmost Atlantic hurricane on record (Lui, 2017). The coastal damages from this storm alone amounted to about €68 million according to initial insurance industry estimates. This includes insured losses and the broader economic impact on businesses and infrastructure. Insurance companies faced €35 million for claims around damage to property, vehicle and assets. Agricultural losses amounted to €10 million in crop, livestock and farm infrastructure damage (Towey, 2018). Coastal areas faced heavy erosion and damage due to high waves and storm surges while beaches, dunes and coastal infrastructure were particularly affected (Ophelia Report, 2017). Storm Ophelia heightened awareness around the risks posed by severe weather events and the need for better preparedness and resilience planning. Over time, specific storms will affect areas of the coast differently, but some areas will persistently be more vulnerable in the long term. This relates to one of the main objectives of this paper: identifying these relatively vulnerable areas.

In a broader context, Ireland is not alone in its heavy placement of location centres on the coastline. According to estimates by (reference), 1 billion people globally reside on land less than 10 metres above current high tide levels, with 230 million living on land less than 1 meter above high tide levels (reference). Over the last century, high sea levels have posed significant risks resulting in the loss of over 8 thousand lives annually and disrupting/displacing approximately 1.5 million people each year on average (reference).

There is also serious economic impact to account for. Mean global flood losses for major coastal cities amount to ~$6 billion per year (reference). Projections by (reference) suggest that this cost may rise to $60billion annually by 2050, even with constant flood probability maintained through adaptations. Without adequate adaptation measures, up to 4.6% of the global population is expected to face annual flooding by 2100, with potential losses reaching 9.3% of global GDP (Reference). Infrastructure vulnerability is another concern, with coastal areas being at an elevated risk of disruptions to power supply cuts or transportation obstructions. Furthermore, environmental risks include the erosion of sandy beaches and saltwater intrusion into agricultural land (Oppenheimer et al., 2019). The disruption of coastal habitats and biodiversity is a longer-term risk at play. (Reference)

Effective adaptation and management strategies can be determined by improving the understanding of the causes and related impacts of sea level extremes, including their role in coastal flooding and erosion.

   1.2 Objectives

[Clearly state the objectives of the report, outlining what you aim to achieve and the questions you seek to answer.]

This work focuses on characterising sea-level extremes using extreme value analysis. This type of analysis utilises a Generalised Extreme Values (GEV) distributions to characterise the tail ends, or extremes, of the overall data distribution. In addition, we consider temporal clustering of the exceedances of return levels calculated from the GEV model. The input data for the GEV model in this analysis is sea-level surge data, which strongly influence coastal flooding and erosion. Estimated return levels based on the data can then be extracted from the model. A return level is a threshold that a particular environmental variable, in this case sea-level surge, is expected to exceed once in a given period of time, such as once in 10, 50, or 100 years. It quantifies the magnitude of extreme events that are likely to occur over different time intervals. Here we focus on the 2-year, 5-year, 20-year, and 100-year return levels for the data, which can be more easily understood as the levels at which we could identify an event as a 1-in-2, 1-in-5, 1-in 20, or 1-in-100-year event.

After this, a plot a temporal clustering of the exceedances of the 1-in-2 and 1-in-5 levels for each station in each year will highlight areas of the coast which face higher counts of spikes above return levels. The objective of clustering the return level exceedances is to identify areas of higher vulnerability, which may be more threatened more often by large spikes in sea level surges.

A graph of a normal distribution

Description automatically generated with medium confidence

*Figure 1: This plot is distinguishing between observed values and expected values. The non-tidal residuals, which are the difference between observed and expected values, represent surges attributed to factors unrelated to tides.*

The variable to be analysed, sea level surges, shown in Figure 1, is driven by storm clusters (or storm sequences), where multiple extratropical cyclones pass through the same location in a given time period (Dacre and Pinto, 2020). This is a basic assumption of our analysis, and a validation against Met Eireann major weather events should show that the exceedances are aligning with storm events.

2. Data

   2.1 Data Sources

[Describe the sources of data used in your analysis. Specify the methodology of data collection if known, and any relevant limitations or biases.]

The sea surge data used in this analysis is sourced from the Copernicus Climate Change Service (C3S), specifically the dataset titled “Global sea level change time series from 1950 to 2050 derived from reanalysis and high resolution CMIP6 climate projections.” This dataset provides a comprehensive time series dataset of global sea level-related variables including tides, storm surges and sea level rise. The data spans from 1950 to 2050, offering insights into historical trends and projections into the future. For this project we focus on the historical and ERA5 reanalysis data which spans from 1979-2014.

**Methodology of Data Collection**

This dataset is generated using the Deltares Global Tide and Surge Model (GTSM) version 3.0. This hydrodynamic model simulates water levels at 10-minute intervals, then it forces input from both reanalysis data and climate models (Muis et al., 2020). Key methodologies included:

* Hydrodynamic Modelling: the GTSM dynamically simulate water levels. It integrates multiple factors such as celestial tide-generating forces and meteorological conditions (wind and pressure at mean sea level).
* Climate Forcing: The historical period is from 1950-2014, the model used climate forcing data from the ERA5 reanalysis and historical simulations from five different Global Climate Models, or GCMs, within the CMIP6 dataset.
* Variable Coverage: The dataset covers scales with grid points at 0.1° resolution across the coastlines. There are also ocean grid points at 0.25°, 0.5° and 1° resolutions, depending on the distance from the coastline.

The main variables of focus in the dataset are:

* Mean Sea Level: The annual mean sea level relative to the 1986-2005 reference period.
* Storm Surge Residual: This calculates as the difference between the total water level and the tidal elevation.
* Tidal Elevation: Derived from GTSM simulations by using only the celestial tide-generating factors.
* Total Water Level: Summed together using contributions from pure tide, storm surge, and changes in annual mean sea level.

There are some relevant limitations and potential biases which can be inferred based on the nature of the dataset. HighResMIP multi-model ensemble is used to quantify uncertainties in the data. This suggests that variability in the projections exists due to differences between the GCMs used. There is also the exclusion of tectonic and subsidence factors mentioned in the description of the mean sea level variable. The temporal coverage of different time spans may affect the continuity and comparability of the dataset (ERA5 reanalysis: 1979-2018, historical climate projections 1950-2014, and future projections: 2015-2050).

The use of the GTSM is preferred in this paper due to a strong coverage in comparison to tidal gauge data around the coast of Ireland. Muis et al. (2020) shows several validations of the model and reasserts its accuracy in mapping sea level surges.

**2.3 Exploratory Data Analysis**

**[Show EDA which helped you gain insights and understand the characteristics of the data. Describe the exploratory analysis techniques used, such as summary statistics, data visualization, or data transformations. Present any interesting patterns, trends, or relationships observed in the data.]**

In this section, we consider an exploratory data analysis. This process is crucial in the understanding of the dataset’s characteristics, identifying patterns, trends and relationships, as well as providing insights that guide further analysis.

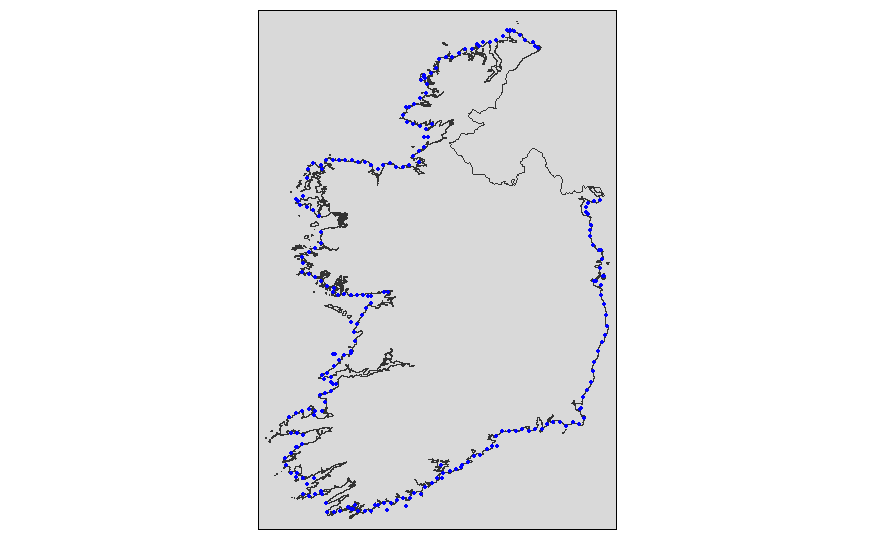
The dataset contains several factors, including the source station name (as a character variable), the year, month, and day of the measurement (as numeric variables), the daily maximum surge value (a numeric variable), the date and time of the measurement (as a POSIXct variable), and the longitude and latitude of the station (as numeric variables). No null or duplicate values exist in the data.

To better facilitate analysis, the *date\_time* column was converted to a date format, and additional time-based features such as year and month were extracted. Furthermore, a custom order for months from July to June was defined to better reflect seasonal patterns, as storm surges tend to spike in the winter months.

Summaries and **Visualisation**

A visualisation of all the stations in the dataset was necessary to show the widespread scope of the coverage of the Irish coastline provided by the dataset. Figures 1 and 2 display, respectively, the locations of all stations in the analysis, and the primary station of focus for the exploratory analysis, Dun Laoghaire.

A map of ireland with a blue dot

Description automatically generated

*Fig 2: All Stations* *Fig 3: Dun Laoghaire Station (795)*

To localise the preliminary exploratory analysis, Dun Laoghaire Station (station ID 795) serves as a representative example. The geographical distribution of all stations around the coast of Ireland is depicted in Figure 2. These visuals provide an overview of the spatial coverage of the dataset. Figure 3 shows only the Dun Laoghaire Station, highlighting its specific location on the map.

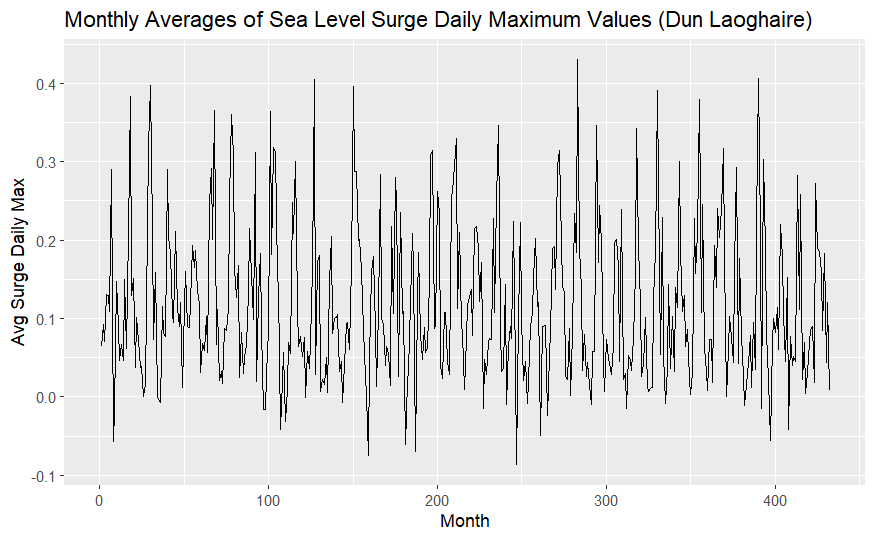
Summary statistics were calculated for the variable of interest, surge\_daily\_max in Dun Laoghaire:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| -0.60200 | -0.03500 | 0.06000 | 0.09162 | 0.18400 | 2.85200 |

*Table A: Summary Statistics for surge\_daily\_max for station Dun Laoghaire.*

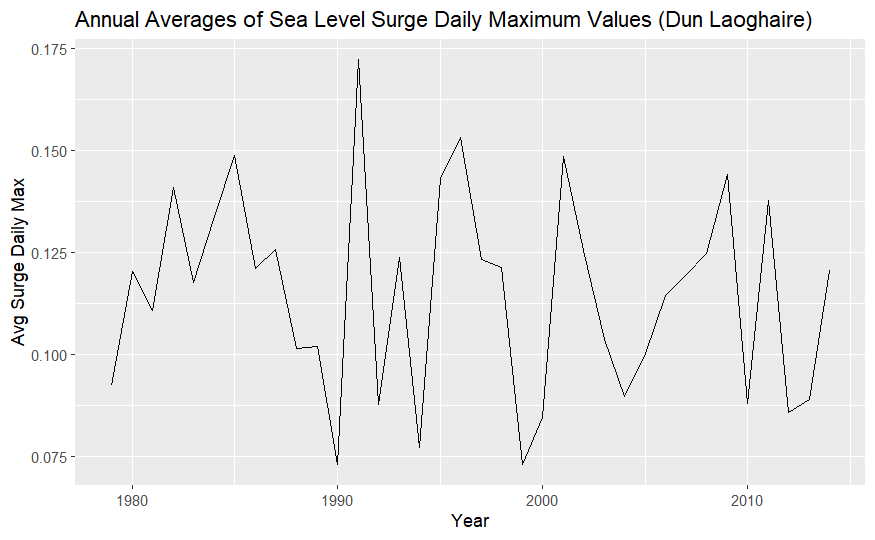
Table A presents the minimum, first quartile, median, mean, third quartile, and maximum values of the sea level surge daily maximum measurements.

The mean surge value (0.09162 meters) is slightly higher than the median (0.060 meters), suggesting a right-skewed distribution where higher surges might be more frequent or extreme. These values lie close to zero because we expect the observed vs expected values to be very close except for in the case of extreme events. In addition, the maximum value is a lot higher than the 3rd Quartile, which highlights a presence of high outliers and extreme events. This is the area we will delve deeper into.



*Figure 4: Monthly Averages of Sea Level Surge Daily Maximum Values for Dun Laoghaire. The x-axis represents the months, while the y-axis shows the average surge values.*

Figure 4 illustrates the monthly averages of sea level surge daily maximum values for Dun Laoghaire. This plot helps in identifying seasonal patterns and trends over time. The cyclical nature of the data shows the expected recurring seasonal variations, highlighting periods of higher and lower surge activity throughout the year.



*Figure 5: Annual Averages of Sea Level Surge Daily Maximum Values for Dun Laoghaire Station. The x-axis represents the years, and the y-axis shows the average annual surge values.*

Figure 5 presents the annual averages of sea level surge daily maximum values for Dun Laoghaire Station over the timeframe of the dataset. This plot highlights long-term trends and potential changes in sea level surges over the years. The variability in the plot indicates fluctuating surge levels, with certain years experiencing higher averages than others, suggesting periods of increased surge activity.

**3. Methods**

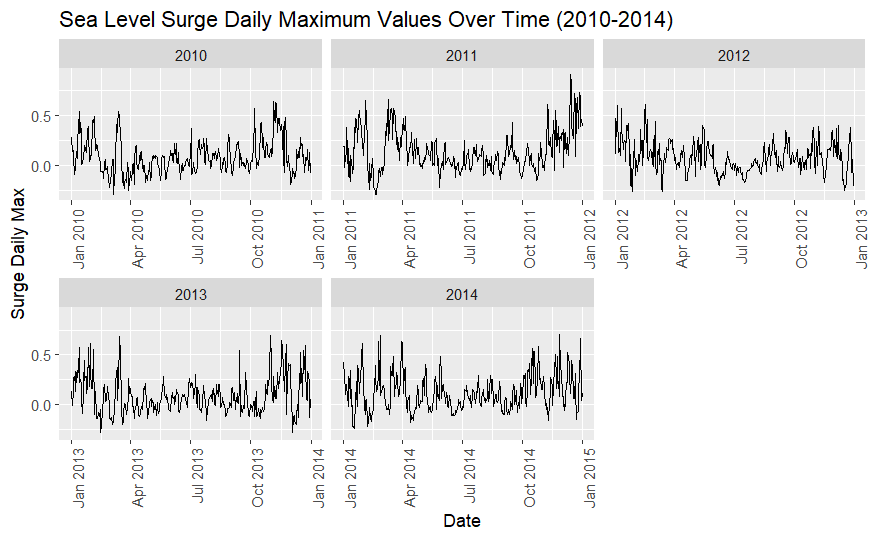
**[Outline the specific methodology or approach used to address the research questions. Describe the statistical models mathematically making sure to explain notation. If exploring different model types, you can include them here and then show the comparison in the results section]**

The first main objective of this paper is to model the extreme values of sea level surges using the Generalized Extreme Value (GEV) distribution. The GEV distribution is suitable for modelling the maximum or minimum of a dataset and is defined by the following parameters: location (𝜇), scale (𝜎), and shape (𝜉). In this case, we model the maxima. The GEV distribution encompasses three types of extreme distributions: the Gumbel, Fréchet and Weibull distributions. Each is used under different circumstances for modelling extreme values. The GEV can decide which of these distributions is most suitable and apply it. The notation for the model is:

ymax (s, t) | μ (s), σ(s), ξ(s) ∼ GEV (μ(s), σ(s), ξ(s))

In this equation, 𝑦max (𝑠, 𝑡) represents the maximum surge at location 𝑠 and time 𝑡, and 𝜇 (𝑠) is the location parameter that varies over space. Location (𝜇) shifts the distribution along the x axis, determining the central tendency of the maxima. (*s*) cause variation in (𝜇) allowing it to capture spatial factors. Scale (𝜎(s)) controls the spread or dispersion of the distribution, and shape (𝜉(s)) defines the tail behaviour, usually taking the form of Gumbel, Fréchet or Weibull distribution.

In practical terms, applying this distribution to our sea level surge data involves estimating parameters 𝜇 (𝑠), 𝜎(s), and 𝜉(s). We can consider the spatial variability by allowing the parameter to vary over space. This allows us to calculate return levels for each location.





*Figure 6: Overview of the sea level surge daily maxima for Dun Laoghaire station from 2010-2014, a subset for clarity purposes. Highlighted in red are the chosen block maxima results.*

The analysis will consist of two different approaches of using a GEV distribution. Firstly, a block maxima approach considers the maximum value for given station each year as the extreme value (as shown in Figure 6). Each maximum point is used to populate the “extreme” value subset from which we fit the GEV distribution. The visualisation of the block maxima is as shown above.

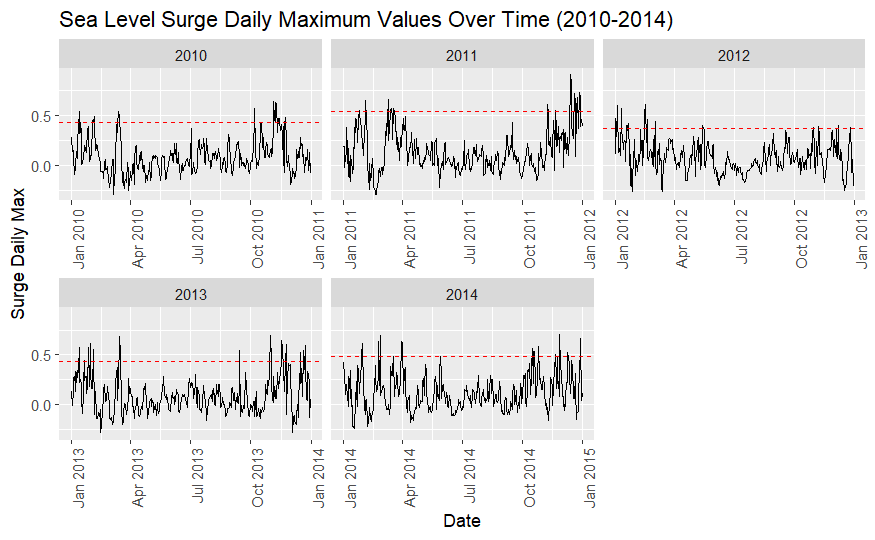


Figure 7: *Overview of the sea level surge daily maxima for Dun Laoghaire station from 2010-2014, a subset for clarity purposes. Values above the dashed line are the chosen threshold results.*

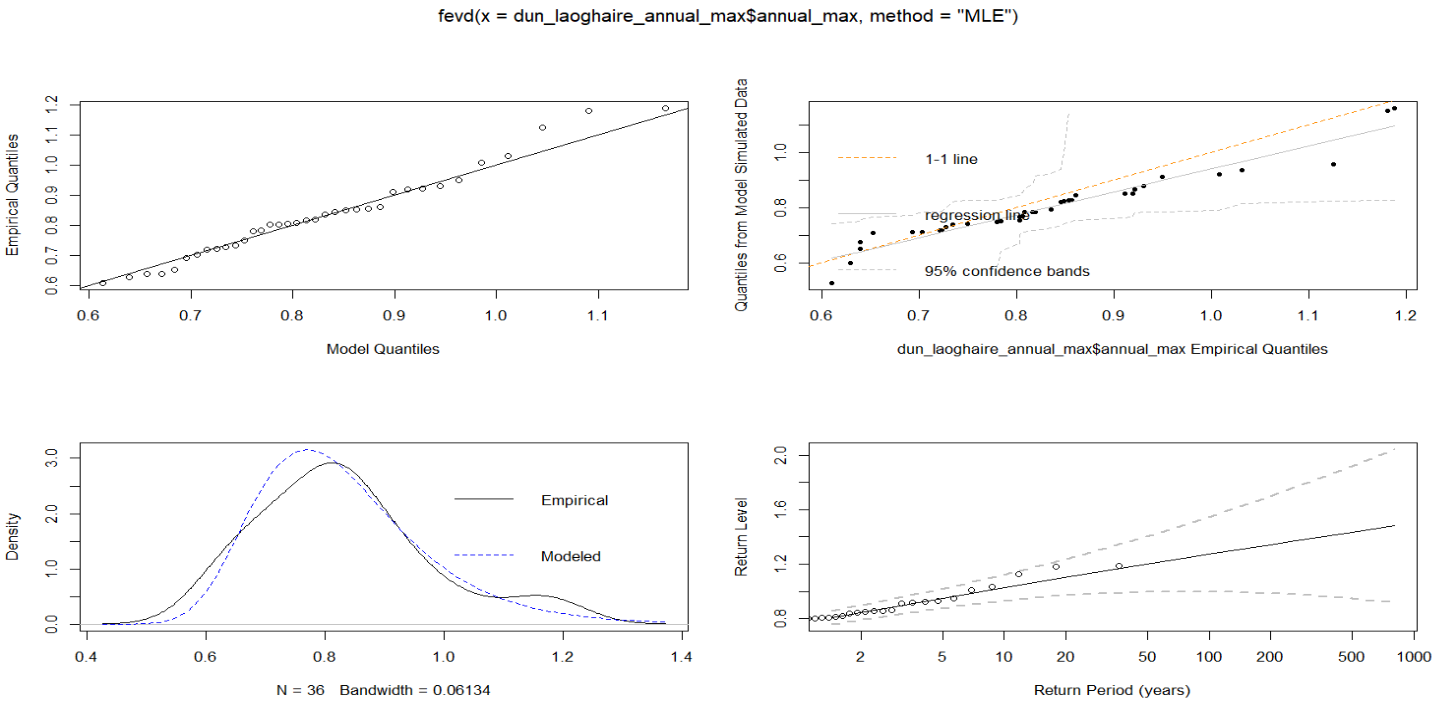
By comparison, a threshold analysis, or Peaks Over Threshold (POT) method, takes a given quantile of the data to use as the extreme subset. For this analysis, the 95th percentile is used as a threshold. Any points in the top 5% of values in each year will be considered to calculate the return levels for a given station. This method is highlighted above in *Figure 7*. According to the Fisher-Tippett-Gnedenko theorem, the block maxima converge to one of the Generalized Extreme Value (GEV) distributions: Gumbel, Fréchet, or Weibull. However, for a POT method, the Generalised Pareto Distribution, and specifically the Pareto distribution, are more adept at capturing the tail behaviour of the distribution. The convergence to the GPD, and thereby the Pareto distribution, when modelling exceedances over a high threshold is mathematically justified by the Pickands-Balkema-de Haan theorem (Pickands, 1975).

Both the block maxima and threshold approaches are to be tested, and their resulting return levels will be compared.

**4. Results**

**[Present the findings of your analysis in a clear and concise manner. Include tables, graphs, or charts as necessary to illustrate the results.]**

***Block Maxima Model Fitting***



*Figure 8: Forecast Error Variance Decomposition (FEVD) for extreme value indices. The four-panel plot illustrates how the contribution of each index to the forecast error variance changes across different time horizons. Each panel represents the decomposition for one extreme value index, showing its relative impact on the forecast error variance over time.*

We are fitting the data using the ‘fevd’ function from the ‘extRemes’ package in R. Figure 8 shows four panels relating fitting the data to a GEV distribution using the block maxima approach. To reiterate, the following results are for our case study location of Dun Laoghaire, however similar results can be seen across all locations.

The top left plot shows the resulting Q-Q plot after fitting the model. Comparing the quantiles of the observed data versus the quantiles of the GEV model assume a good fit, with most points lying close to the 45-degree line.

The probability plot in the top right assesses the goodness of fit with the x-axis representing the empirical quantiles of the observed data and the y-axis demonstrating the quantiles from the model’s simulated data. Most points are on or very close to the 1-1-line, indicative of a near-perfect fit.

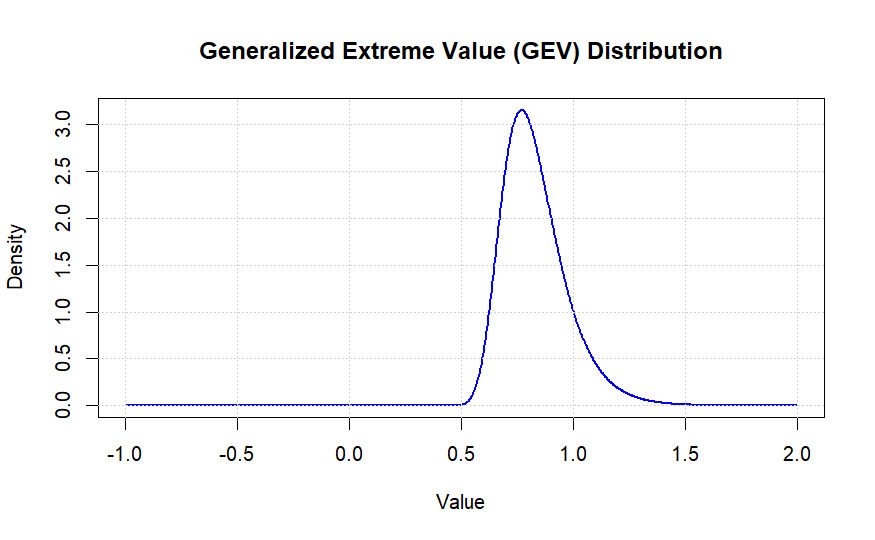
The density plot compares the density of the observed data with the density implied by the model. The empirical and modelled lines show significant overlap which implies good fit.

Finally, the return level plot (bottom right) shows estimated return levels for different return periods based on the GEV fitted model. The x-axis shows the logarithmic return period (years) and the y-axis shows the return level for which storm surge height is expected to be exceeded once every given period of return. Return level estimates are all inside the 95% confidence interval (the dashed lines) which again shows a strong reliability for estimating extreme sea level surges.

***Block Maxima Results***

Firstly, the GEV is fitted to the data associated with station 795. This provides a singular point of comparison in assessing the strength of fit of the model. The Generalized Extreme Value (GEV) distribution parameters using the block maxima approach were estimated using the Maximum Likelihood Estimation (MLE) method. The negative log-likelihood value for the model fit was -20.97111, indicating reasonable fit to the annual maximum surge data.

The estimated parameters for the GEV distribution are as follows: location (μ) = 0.7659, scale (σ) = 0.1167, and shape (ξ) = -0.0274. The standard errors for these estimates are 0.0222, 0.0162, and 0.1354 respectively, suggesting acceptable precision in the parameter estimates.



*Figure 9: Block maxima extreme value distribution for station 795 (Dun Loaghaire).*

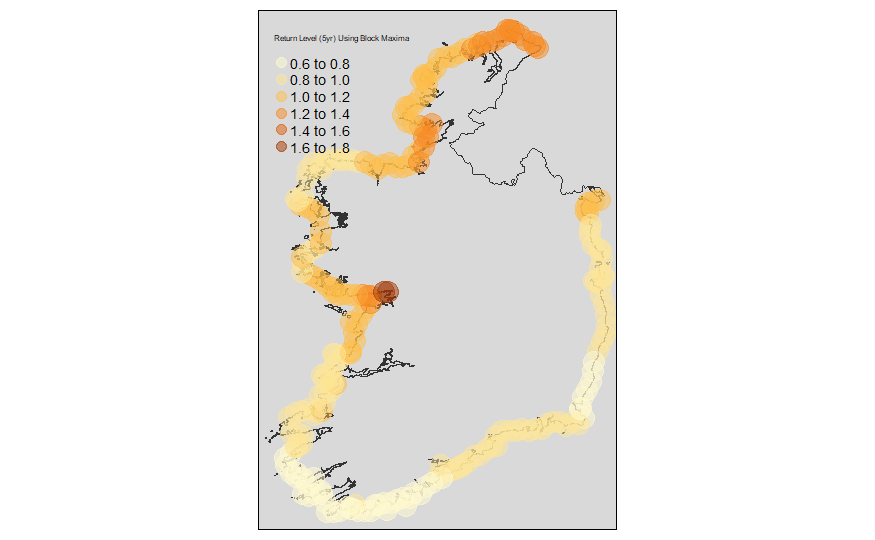
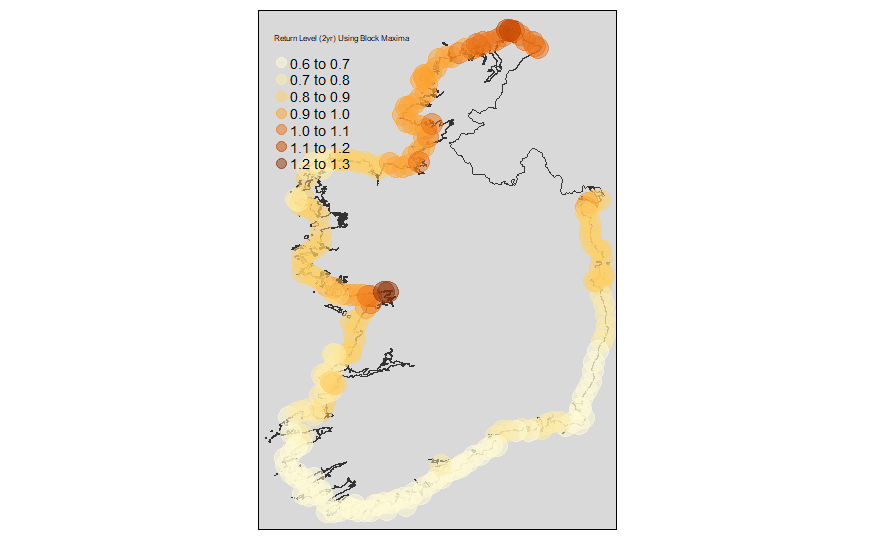
Figure 9 shows the resulting distribution of the extreme values for station 795 (Dun Laoghaire). The shape parameter being very close to zero suggests a Gumbel distribution.

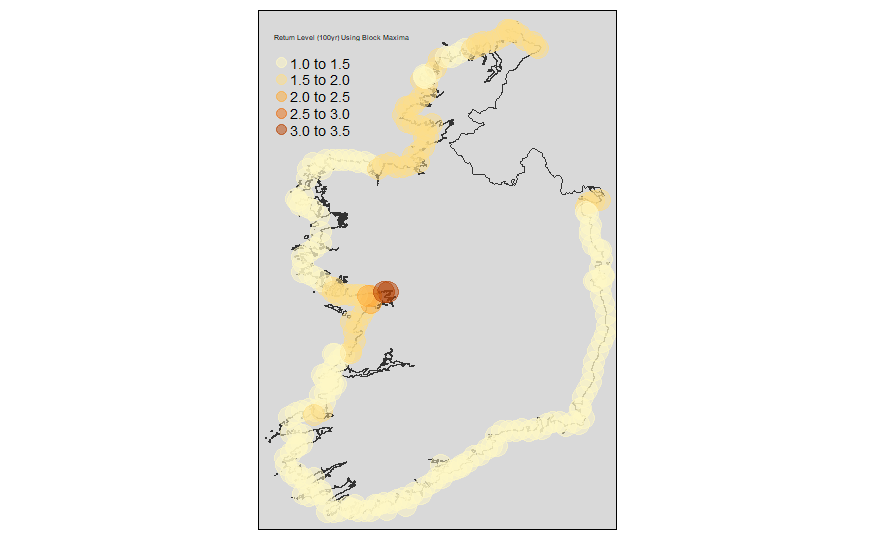
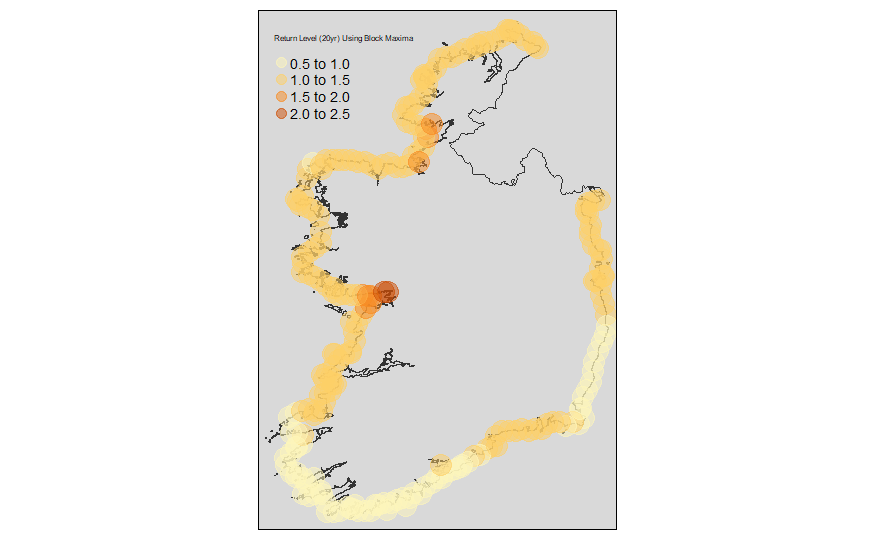
The covariance matrix of the parameter estimates is shown in Table X (Appendix). This matrix provides insights into the relationships between the estimated parameters, with off-diagonal elements indicating the covariances.

The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values for the model are -35.94222 and -31.19166, respectively. These values support the adequacy of the model for the given data. (Will I not use these to compare with the threshold results?)

For station 795, the return levels for 2-year, 5-year, 20-year, and 100-year periods were estimated as 0.8085, 0.9374, 1.0989, and 1.2704 meters, respectively. The 95% confidence intervals for these return levels provide a range within which the true values are expected to lie, indicating the uncertainty in these predictions (Table Y - Appendix).

The GEV model was applied to all the stations in the dataset. The resulting return levels for the 2, 5, 20 and 100-year return levels were then plotted on a heatmap of the Irish coast as shown in figure 10 below.

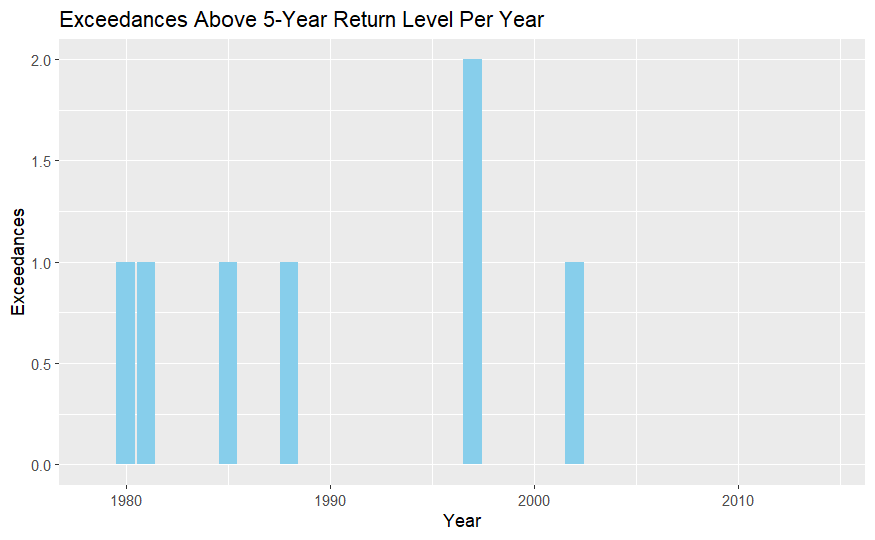




*Figure 10: Block maxima calculated return levels mapped continuously for the 2-, 5-, 20-, and 100-year periods respectively.*

Figure 10 shows the block maxima calculated return levels around the coast for 2-, 5-, 20-, and 100-year periods respectively. Darker-coloured areas indicate higher return levels while lighter colours have lower return levels.

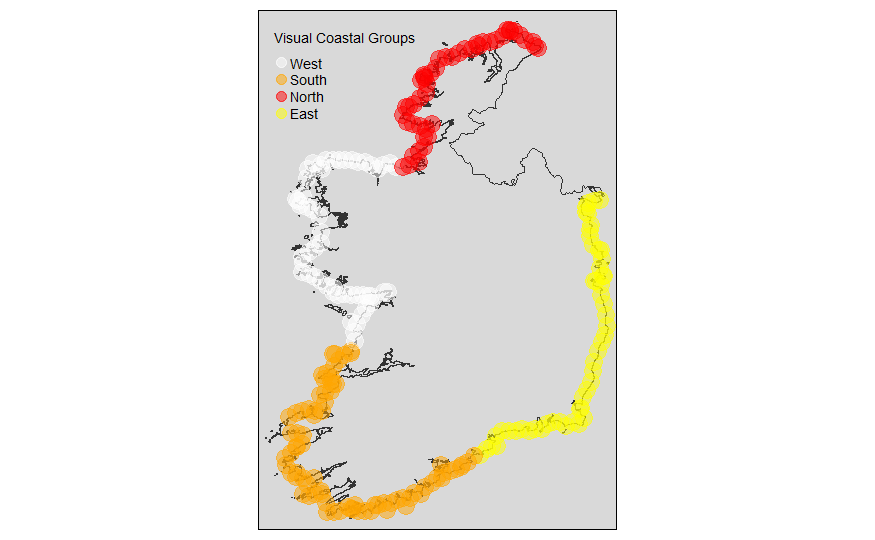
The next objective was to calculate the exceedances of the 5-year return level for each station. Again, station 795 is used initially to plot a singular location’s exceedance count over time.



*Figure 11: Dun Laoghaire exceedance counts of 5-yr return level (Block Maxima).*

Figure 11 shows the exceedance counts where sea surge spiked above the 5-year return level in Dun Laoghaire station 795, for each year.

Before looping this process to count exceedances of the 5-year level, it is important to cluster our locations to more easily draw conclusions and meaning from the final exceedance analysis. Location clustering was performed to group stations into different coastal groups: North, East, South, and West.

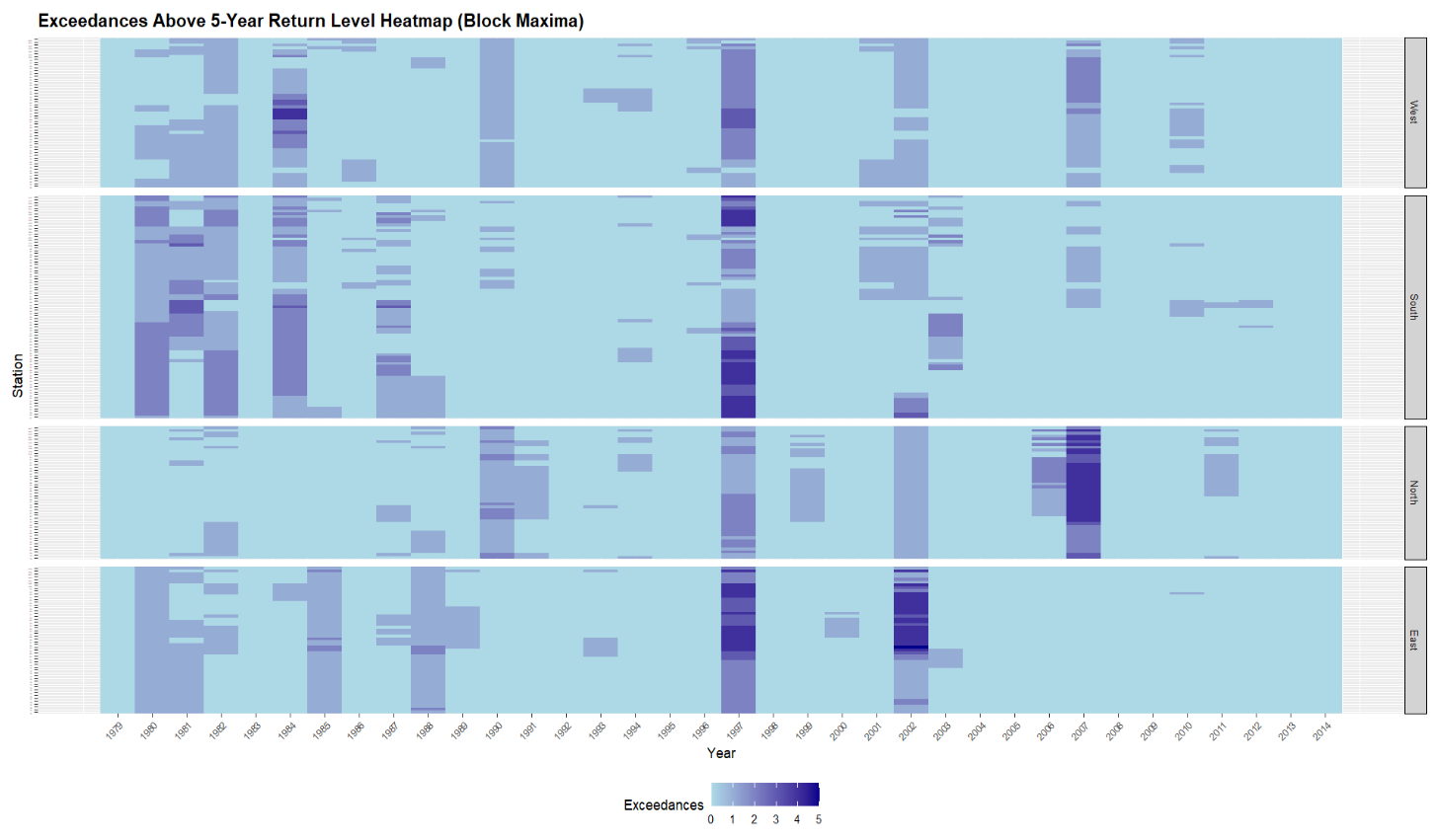


*Figure 12: Coastal grouping resulting from a k=4 location clustering using k-means.*

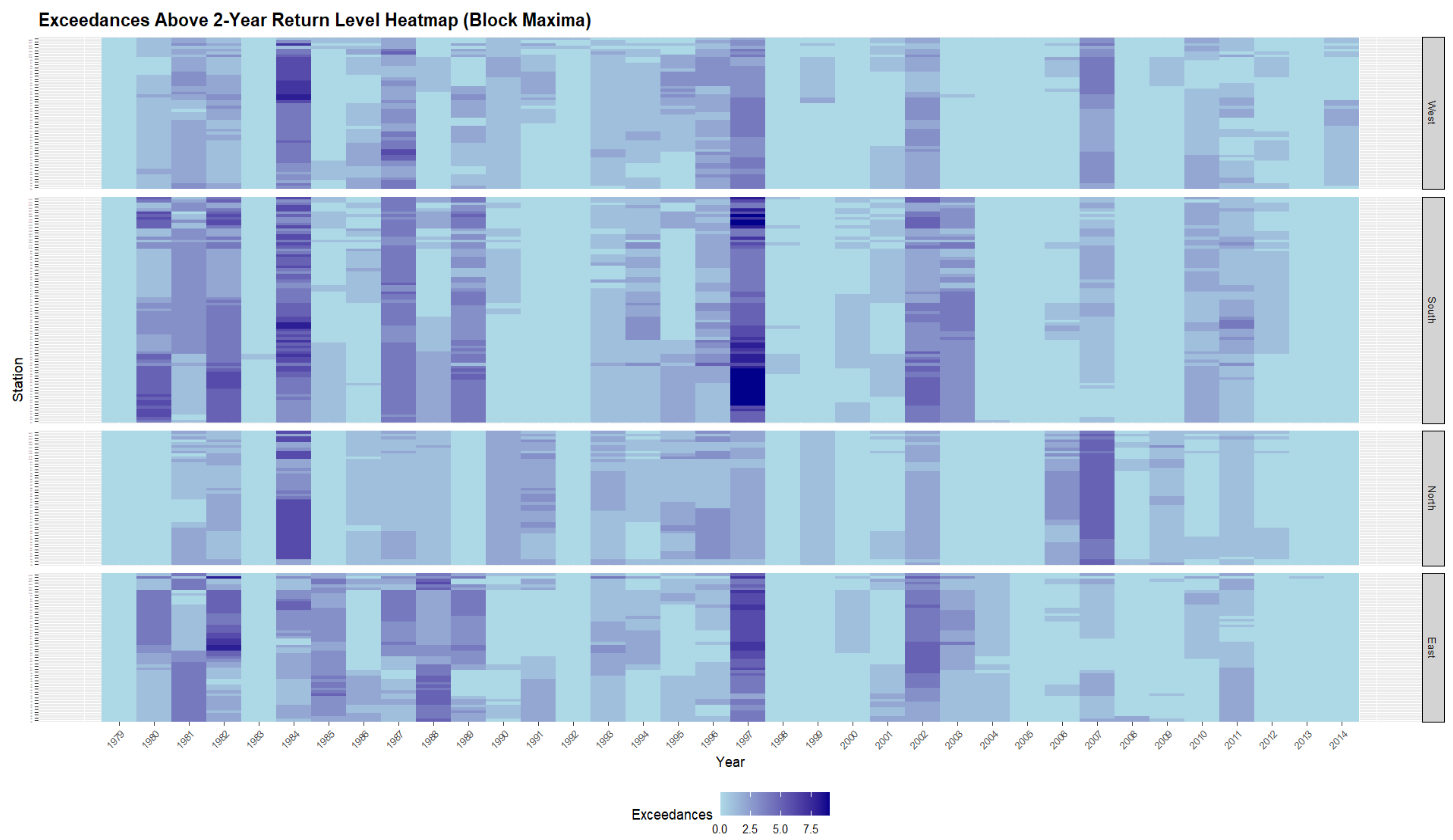
Figure 12 shows the results of the K-means clustering. Four major coastal regions are identified and labelled according to their approximate facings: North, West, South and East.

Note that some stations in Galway are grouped with the Northern cluster. This does not cause much issue in our analysis as the clustering of coastal regions serves solely to provide greater context and approximate grouping for the temporal clustering of exceedances, and a small number of outlier stations does not affect the overall analysis.

The next step is to map the exceedances of the 5-year and 2-year return levels across all stations and group/cluster them based on the North, South, East, West grouping. This will give us a visual of areas within close proximity which experience exceedances above the 5/2-year level in a given year.



*Figure 13: Temporal clustering heatmap of exceedances of the 5-year return level by each station over time, grouped by coast.*

 *Figure 14: Temporal clustering heatmap of exceedances of the 2-year return level by each station over time, grouped by coast.*

Figures 13 and 14 shows a temporal clustering analysis heatmap of exceedances above the 5-year and 2 -year return levels. Grouped by coastal regions, the y-axis represents the coastal stations. The x-axis shows the years of the dataset from 1979-2014. Darker areas of the heatmap indicate higher exceedances in the given year for that given station.

|  |  |  |  |
| --- | --- | --- | --- |
| Cluster | Exceedance Count (5yr) | Station Count (5yr) | Exceedances per Station (5yr) |
| West | 509 | 53 | 9.60 |
| South | 886 | 79 | 11.22 |
| North | 490 | 47 | 10.43 |
| East | 549 | 52 | 10.56 |

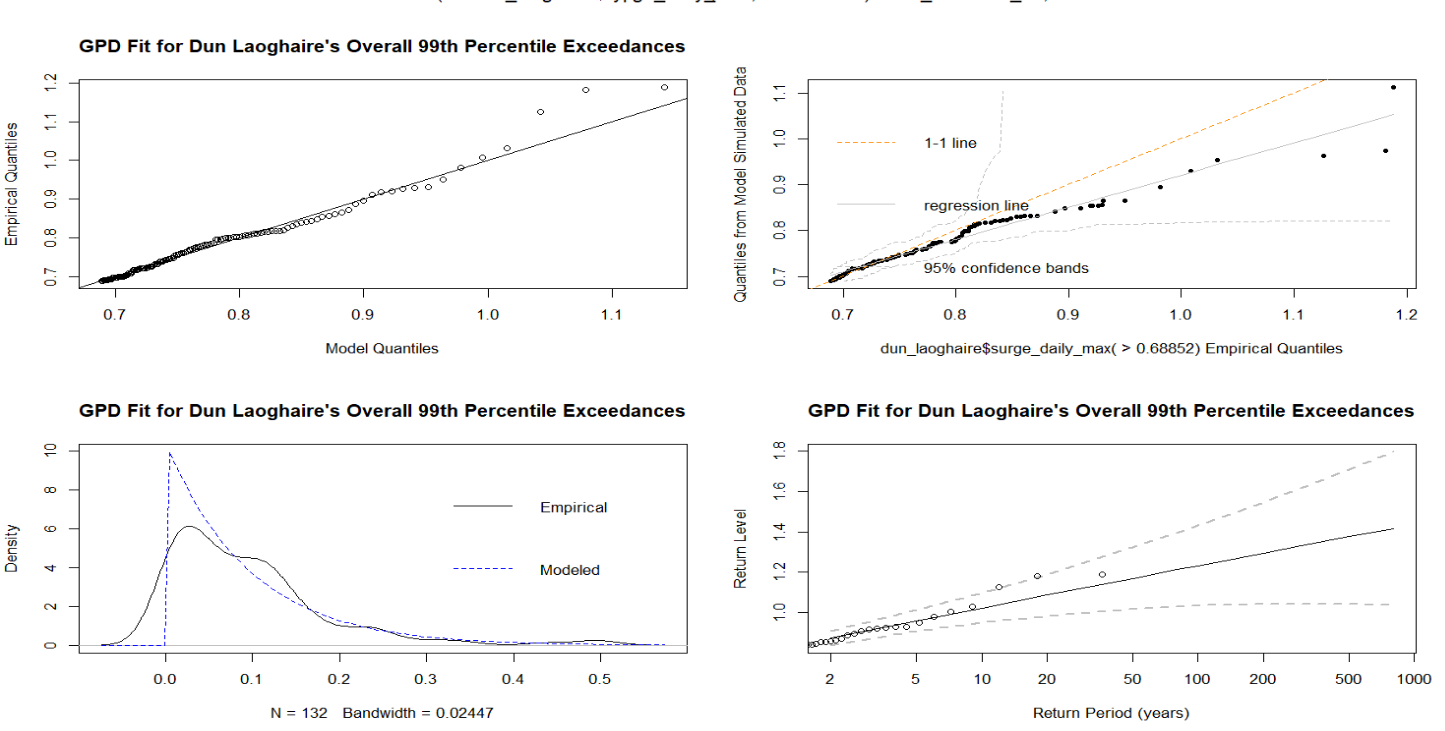
*Table B: Average exceedances by cluster group of 5-year return level using a Block Maxima method.*

|  |  |  |  |
| --- | --- | --- | --- |
| Cluster | Exceedance Count  (2yr) | Station Count (2yr) | Exceedances per Station (2yr) |
| West | 1801 | 53 | 33.98 |
| South | 3712 | 79 | 46.98 |
| North | 1641 | 47 | 34.91 |
| East | 2210 | 52 | 42.50 |

*Table C: Average exceedances by cluster groups of 2-year return level using a Block Maxima method.*

Tables B and C show the cluster groups, their respective exceedance counts, station counts, and the exceedances per station, for the 5-year and 2-year return levels. By 5-year returns, the South cluster is the highest in exceedances per station with 11.22, and a difference of 0.66 exceedances to the East cluster in second. Closely following is the North, then the West cluster yielding the least exceedances per station. At a 2-year return level, the South has a larger lead in exceedances with 46.98 exceedances, followed by East, North, and West. The results of the block maxima indicate a tendency for the South to be more vulnerable to spikes above 2-year and 5-year levels relative to the rest of the coastline.

***Peaks-Over-Threshold (POT) Model Fitting***



*Figure 15: Forecast Error Variance Decomposition (FEVD) for extreme value indices. The four-panel plot illustrates how the contribution of each index to the forecast error variance changes across different time horizons. Each panel represents the decomposition for one extreme value index, showing its relative impact on the forecast error variance over time.*

We are fitting the data using the ‘fevd’ function from the ‘extRemes’ package in R. Figure 15 shows four panels relating fitting the data to a GPD distribution using the POT approach. To reiterate, the following results are for our case study location of Dun Laoghaire, however similar results can be seen across all locations.

The top left plot shows the resulting Q-Q plot after fitting the model. Comparing the quantiles of the observed data versus the quantiles of the GPD model assume a good fit, with most points lying close to the 45-degree line. There may also be a couple of outliers seen in the higher quantiles.

The probability plot in the top right assesses the goodness of fit with the x-axis representing the empirical quantiles of the observed data and the y-axis demonstrating the quantiles from the model’s simulated data. Most points, except a small number of potential outliers, are on or very close to the 1-1-line, indicative of a good fit.

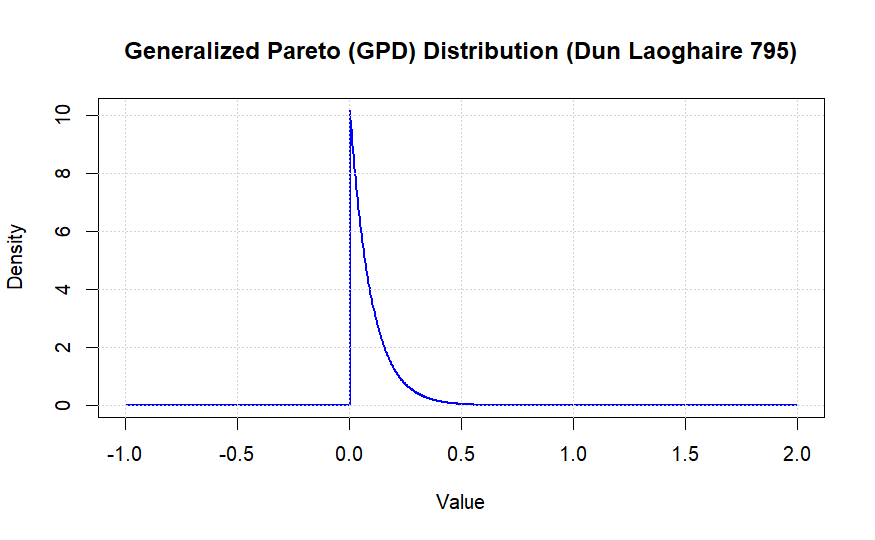
The density plot compares the density of the observed data with the density implied by the model. The empirical and modelled lines do not show a strong correlation. This is because the GPD distribution …... (Not sure what to say here could use some help)

Finally, the return level plot (bottom right) shows estimated return levels for different return periods based on the GPD fitted model. The x-axis shows the logarithmic return period (years) and the y-axis shows the return level for which storm surge height is expected to be exceeded once every given period of return. Return level estimates are all inside the 95% confidence interval (the dashed lines) which again shows a strong reliability for estimating extreme sea level surges.

***Peaks-Over-Threshold Results***

The same methods will be applied to the data in this section using a Peaks-Over-Threshold (POT) approach instead of a block maxima approach. Again, we use station 795 Dun Laoghaire as a comparison location. The key difference in for the POT is that we are fitting a Generalised Pareto Distribution (GPD) to the values above the threshold. The GPD distribution parameters are scale and shape. We do not estimate the location parameter in this case because the location is the threshold itself, which is chosen to be the 99th percentile. The scale and shape parameters are estimated using the Maximum Likelihood Estimation (MLE) method. The negative log-likelihood value for the POT model fit was -180.0479, indicating a significantly lower and therefore better fit to the annual maximum surge data.

The estimated parameters for the GEV distribution are as follows: scale (σ) = 0.095 and shape (ξ) = -0.0104. The standard errors for these estimates are 0.0116 and 0.0857 respectively, suggesting acceptable precision in the parameter estimates.



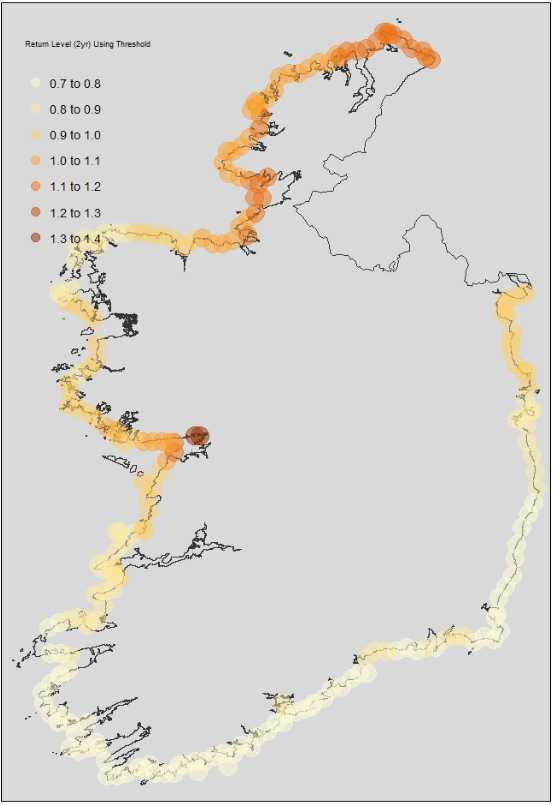
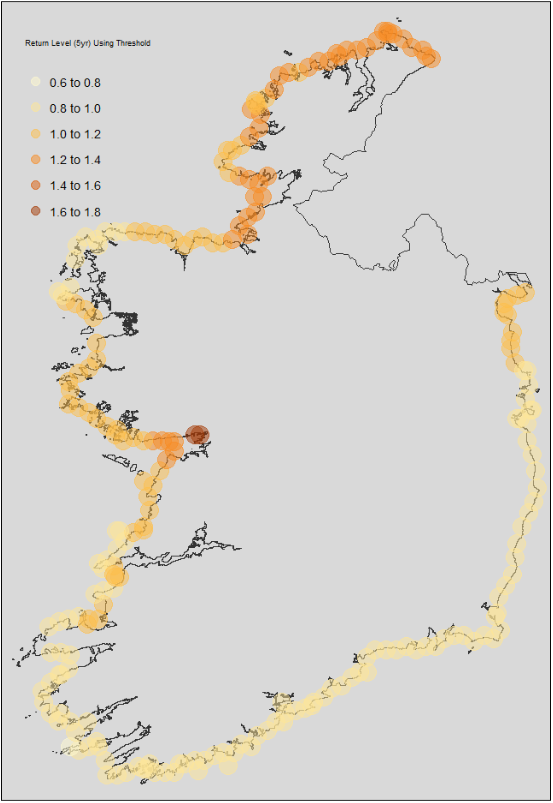
*Figure 16: Generalised Pareto Distribution (POT Method) for station 795 (Dun Loaghaire).*

Figure 16 shows the resulting distribution of the extreme values for station 795 (Dun Laoghaire). The covariance matrix of the parameter estimates is shown in Table V (Appendix) with one less parameter than before, excluding location.

The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values for the model are -356.1 and -350.3, respectively. These values are ten-fold lower than the comparable results in the block maxima approach, supporting strongly the suitability’ of the model for the given data. (Will I not use these to compare with the bm results?)

For station 795, the return levels for 2-year, 5-year, 20-year, and 100-year periods were estimated as 0.8759, 0.9608, 1.0877 and 1.2327 meters, respectively. The 95% confidence intervals for these return levels provide a range within which the true values are expected to lie, indicating the uncertainty in these predictions (Table W - Appendix).

The GPD model was applied to all the stations in the dataset. The resulting return levels for the 2, 5, 20 and 100-year return levels were then plotted on a heatmap of the Irish coast as shown in figure 17 below.

A screenshot of a computer

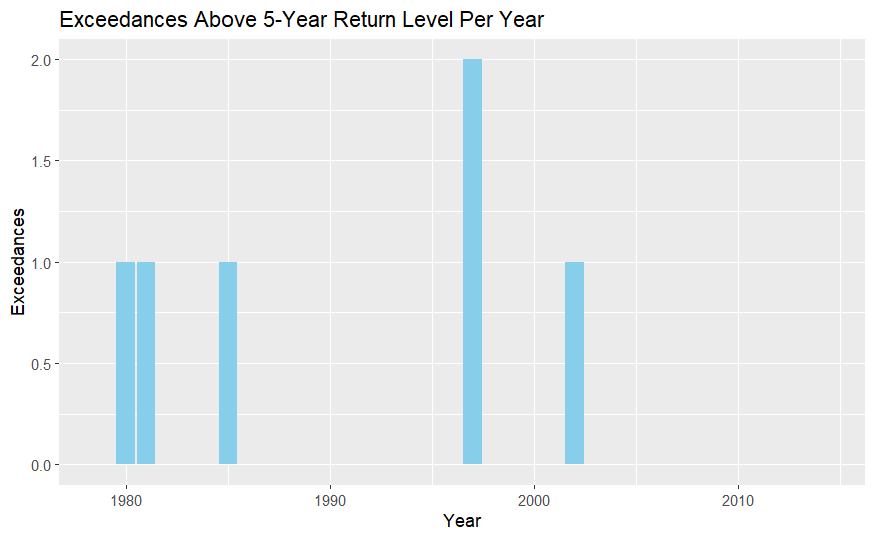
Description automatically generated A map of a country

Description automatically generated

*Figure 17: Peaks-Over-Threshold calculated return levels mapped continuously for the 2-, 5-, 20-, and 100-year periods respectively.*

Figure 17 shows the POT calculated return levels around the coast for 2-, 5-, 20-, and 100-year periods respectively. Darker-coloured areas indicate higher return levels while lighter colours have lower return levels.

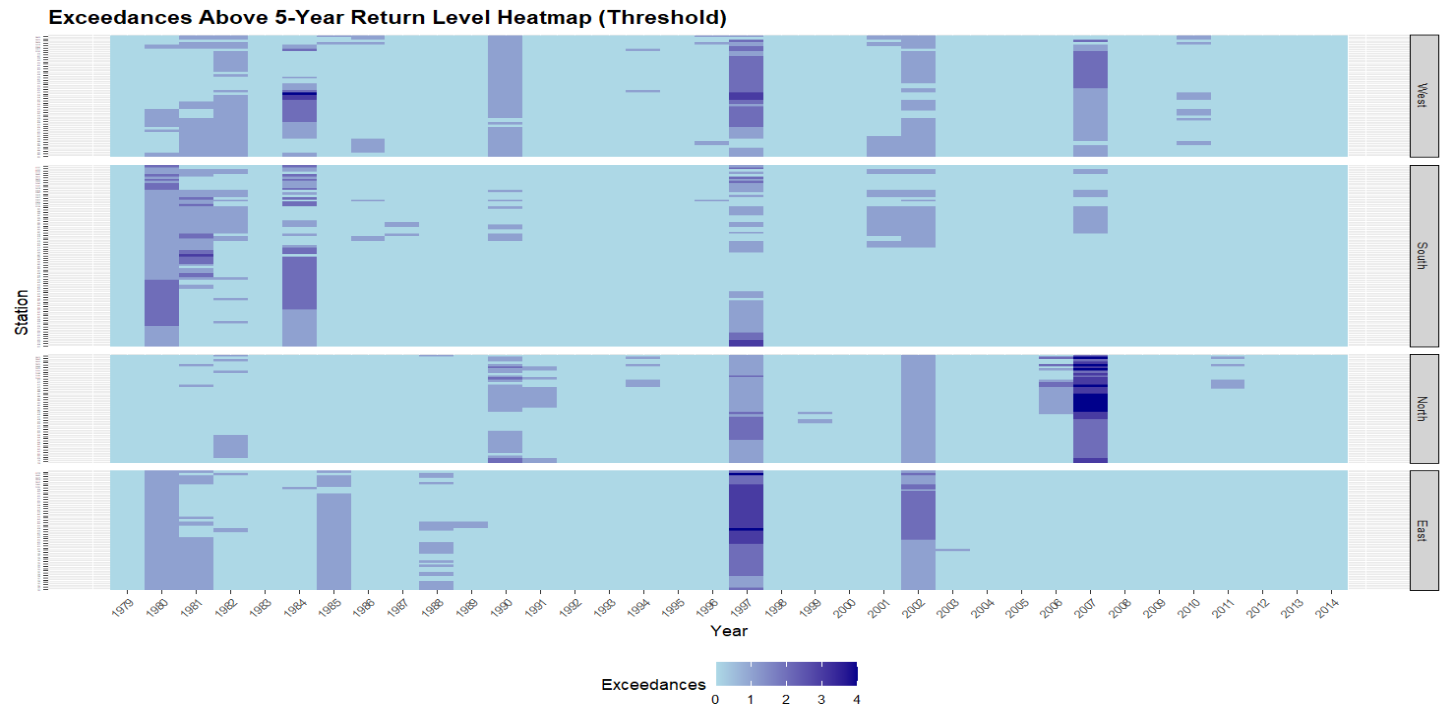
We follow the same order of objectives by calculating the exceedances of the 5-year return level for each station. Again, station 795 is used initially to plot a singular location’s exceedance count over time.



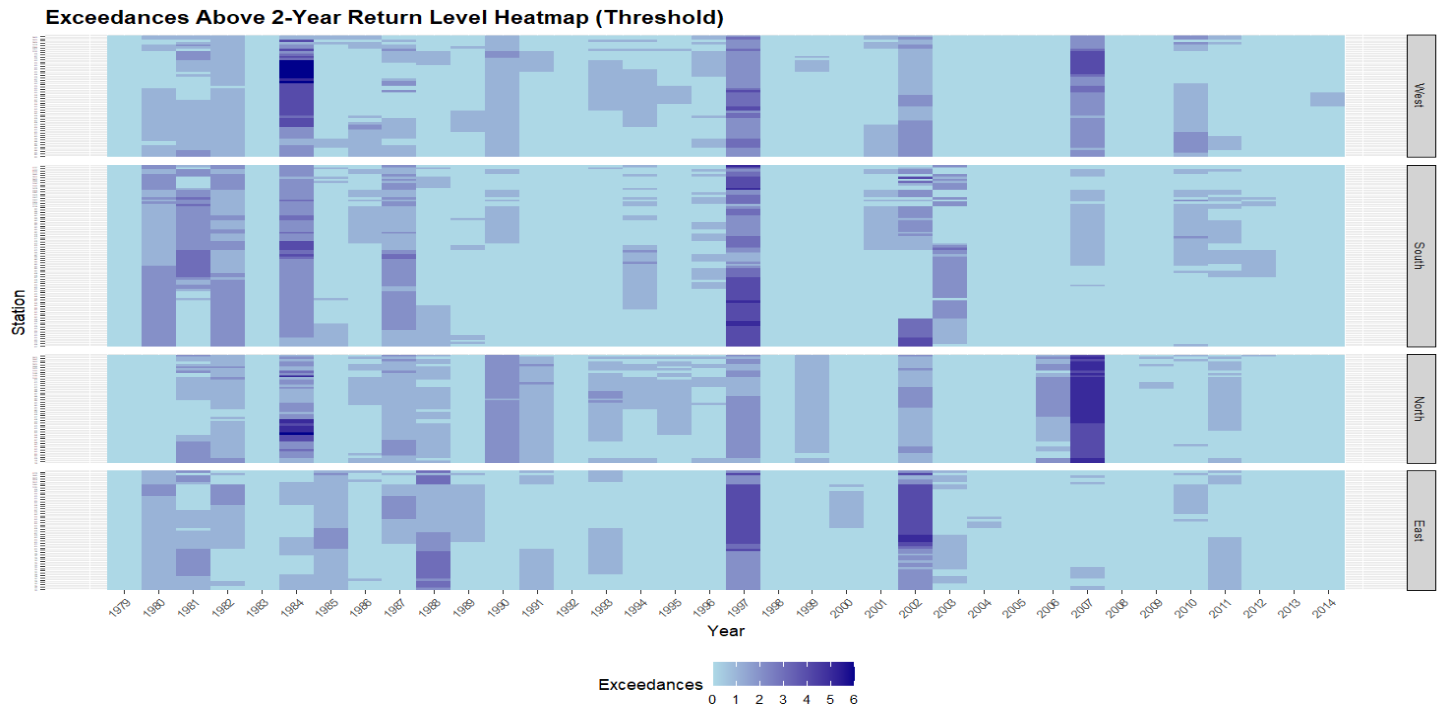
*Figure 18: Dun Laoghaire exceedance counts of 5-yr return level (Peaks-Over-Threshold).*

Figure 18 shows the exceedance counts where sea surge spiked above the 5-year return level in Dun Laoghaire station 795, for each year. This yields similar results to the exceedances for station 795 using block maxima.

The next step is to map the exceedances of the 5-year and 2-year return levels across all stations and group/cluster them based on the North, South, East, West grouping. This will give us a visual of areas within close proximity which experience exceedances above the 5/2-year level in a given year.



*Figure 19: Temporal clustering heatmap of exceedances of the 5-year return level by each station over time, grouped by coast (POT Method).*



*Figure 20: Temporal clustering heatmap of exceedances of the 2-year return level by each station over time, grouped by coast (POT Method).*

Figures 19 and 20 show a temporal clustering analysis heatmap of exceedances above the 5-year and 2 -year return levels. Grouped by coastal regions, the y-axis represents the coastal stations. The x-axis shows the years of the dataset from 1979-2014. Darker areas of the heatmap indicate higher exceedances in the given year for that given station.

We can then compare the results of the block maxima exceedance analysis with the same results using a Peak-Over-Threshold method.

|  |  |  |  |
| --- | --- | --- | --- |
| Cluster | Exceedance Count  (5yr) | Station Count (5yr) | Exceedances per Station (5yr) |
| West | 415 | 53 | 7.83 |
| South | 413 | 79 | 5.23 |
| North | 342 | 47 | 7.28 |
| East | 363 | 52 | 6.98 |

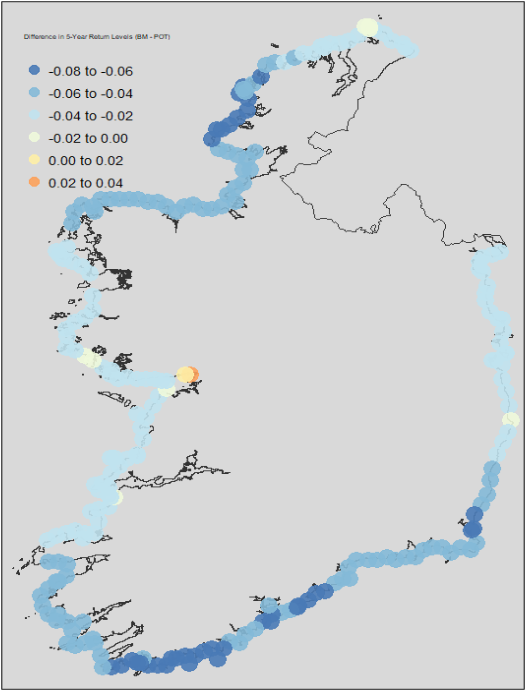
*Table D: Average exceedances by cluster group of 5-year return level using a Peaks-Over-Threshold (POT) Method.*

|  |  |  |  |
| --- | --- | --- | --- |
| Cluster | Exceedance Count  (2yr) | Station Count (2yr) | Exceedances per Station (2yr) |
| West | 961 | 53 | 18.13 |
| South | 1377 | 79 | 17.43 |
| North | 1061 | 47 | 22.57 |
| East | 873 | 52 | 16.79 |

*Table E: Average exceedances by cluster group of 2-year return level using a Peaks-Over-Threshold (POT) Method.*

In the POT approach, we find a difference in the ranking of clusters. The West cluster leads exceedances for the 5-year return levels, while the South cluster this time experiences the lowest exceedances. At a 2-year return, the North cluster has the most exceedances and the East experiences the least. A notable feature of the exceedance results using a POT approach is that there are much less exceedances of both the 5 and 2-year levels.

Finally, it is important to understand how the return levels differ between the Block Maxima and POT approaches. We can map these differences for visual purposes.



*Figure 21: Map showing which locations yield higher 5-year return levels for the POT method vs the BM method.*

Figure 21 shows the actual differences in the calculated 5-year return levels for the Block Maxima method versus the POT method. The differences are shown on a blue/yellow/red continuous colour scale, with negative differences (blue) showing areas with higher estimations in the POT analysis and positive differences (red) showing areas with higher estimations on the Block Maxima.

5. Discussion

[Interpret and discuss the results obtained from the analysis. Explain the implications of the findings, their significance, and any limitations or potential biases in the data or methodology.]

The block maxima extreme analysis produced a valuable temporal clustering heatmap which shows a time and location for specific clustering events of sea surge exceedances above a 5-year return level. In theory, given the close relationship between sea surge and storm surge, this heatmap should reasonably coincide with actual storm events, and show the areas which are vulnerable during those events.

As a validation measure, Ireland’s national meteorological service, Met Eireann’s reporting of major weather events should correlate tightly with the temporal clustering heatmap. Large clustering exceedances appear in 2007 (North), 2002 (East/South), 1997 (East/South), and 1984 (West/North). Met Eireann reports 2 flooding events on the east and south coasts in 2002, as well as extensive flooding in the south and east and a windstorm event in 1997. These two years match our analysis perfectly. However, the large clusters in 2007 and 1984 do not seem to have correlating major event reports by Met Eireann, suggesting a missing link in our analysis.

There are some reasons which may help to explain this discrepancy between the data and the storm reports. Firstly, perhaps sea surge data on its own is not fully predictive of storm events. There may be other influences on sea surges which are not related to storm surges. Astronomical tides can cause extreme sea surges, better known as spring tides or king tides. These occur when the gravitational alignment of the moon, sun and Earth align, causing higher than normal tides (Woo, 2014).

Seiches are standing waves which oscillate in enclosed or partially enclosed bodies of water, such as bays or lakes. Seiches are caused by sudden changes in atmospheric pressure, strong winds or seismic activity (Encyclopedia Britannica). Seiches may affect areas such as Galway Bay or Dublin Bay and may explain why return levels are calculated much higher in Galway Bay, leading it to be grouping outlier in the k-means clustering analysis.

Oceanic and atmospheric patterns, such as the North Atlantic Oscillation (NAO), may be a factor influencing sea surges around the Irish coast. "The North Atlantic Oscillation influences the climate variability in the region and can lead to significant sea level changes along the Irish coast" (Dahlman, 2009).

Another possibility is that the hydrodynamic model, the GTSM, not fully accurate in simulating sea surges around the Irish coastline. The GTSM is widely used for predicting tidal and storm surges. However, like any model, it has limitations. Ireland has a varied and complex coastline. Resolution limitations, or the inability to accurately capture nuances such as coastal topography and bathymetry, small bays, estuaries and coastal shelves, which are prevalent around Ireland, may be causing inaccuracies in the data simulation. This point is made multiple times by Wang et. Al., 2021: “Although the behavior of tides and surges is quite linear for the deep ocean and steep coasts, there may be significant non-linear interaction between tides and surges on the coastal shelf." They also mention that the GTSM uses an unstructured grid to “apply a higher resolution near the coast where the spatial scales are smaller." This acknowledges that coastal areas require higher spatial resolution due to their smaller scales, and implies that without this higher resolution, the model might not capture accurately the coastal dynamics.

Model calibration and validation may play a part too. If the GTSM is not calibrated with local data specific to the Irish coastline it may produce inaccurate results. Stammer et. Al, 2014 outlines the importance in continuous calibration with local observational data, after a review of the accuracy of different global tide models.

Finally, perhaps Met Eireann’s reports are not inclusive of smaller storm events, or more localised events, which may account for the exceedances in the heatmap. A possible issue is that the threshold for what is considered a “reportable” major weather event is too high to capture the 2007 and 1984 exceedance clusters. There may have been storms regardless which disproportionately yielded higher sea surges than a typical storm of the same size. Otherwise, the reports may be national in scope and not register more localised events which cause sea surges.

Regardless of the potential limitations and unknown reasons behind some exceedance clusters, the clustering of exceedances clearly identifies known major weather events accurately. Using the results, we can simply sum the total exceedances for each coastal grouping and divide by the number of stations within each given group to measure the “vulnerability” of each coastal group. This vulnerability is assuming that more spikes above the 2 and 5-year return levels are causing more frequent sea surges above the expected or normal observed levels, which in turn indicates more extreme events compared to other locations. We exclude larger return levels (20 and 100 year) because they do not provide enough exceedances to draw meaningful conclusions from given our 35-year history in this dataset.

**Higher Return Levels with the POT Method**

The POT analysis is calculating higher return levels which is resulting in less exceedances. Understanding why the POT method would return higher levels than the BM method is an important in drawing conclusions from the analysis. There are several reasons which may explain this variability.

The variability may be impacted by the ability of the POT method to capture a higher number of very extreme events which may cause high sea surges multiple times over the course of days. This more detailed view may be accounting for high-magnitude events which would only be counted once in the BM method, which takes only the maximum result.

There may also be statistic modelling differences. The POT method uses a Generalised Pareto Distribution which fits the tail of the data more precisely by using multiple exceedances. This GPD tail is more sensitive to rare, high extremes which can result in higher estimated return levels, especially for longer return periods.

The threshold selection is crucial in the POT analysis. A lower threshold will include more moderate events, while higher thresholds will focus on more extreme values. The threshold used in this case is a 99th percentile, which could be including only the highest data points, causing the return levels to rise.

Ultimately, the variability in extreme values is likely to influence the higher return levels. Regions with high variability or more frequent severe events are captured better by the POT. This will cause the fitted distribution to predict higher return levels as it is more sensitive to extreme data, especially when choosing a high threshold.

**Interpretation of Results**

Given that POT is yielding higher return levels than BM in this analysis, this might indicate several insights about the coastal clusters:

*POT Sensitivity to Frequent Extremes*

The POT leading to higher return levels suggests that there are frequent significant surges in the majority of locations around the coastline, and indicates that for locations where this is true, consistent moderate to extreme events are more likely. This sensitivity means that the POT captures the clustering of extreme events better, which is crucial in understanding the overall exposure and risk.

Figure 21 in the results section illustrates a comparison of the 5-year return levels calculated using both methods. Blue areas show where the 5-year return level was higher for POT than BM, and the red shows vice versa. The blue areas are suggestively more prone to frequent significant sea level surges. The red areas, specifically Galway Bay, by comparison experience less common but more extreme events.

Ultimately, the POT is considering multiple significant surges within the same year which provides a broader view of extreme event frequency. The BM method is sensitive to the single largest event in a year, making it influenced by rare, exceptionally high surges. Blue zones (where 5-yr return levels for POT > BM) have typically more frequent extreme events, while Galway Bay which represents the red zone is prone to rare but extremely significant surges. The blue zones, by experiencing more consistent moderate to extreme surges, are more vulnerable to damage over time, impacting infrastructure, erosion rates and coastal ecosystems. Red zones are more vulnerable to severe extreme sea surges, which have potential for catastrophic events, and should be more prepared against emergency situations with resilient infrastructure in place.

6. Conclusion

[Summarize the key findings of the report and their implications. Restate the objectives and research questions and provide a concise summary of the main results obtained.]

This analysis underscores the importance of using both BM and POT methods to get a comprehensive understanding of coastal vulnerabilities. Each method highlights different aspects of extreme event behavior, which together provide a fuller picture of risk and exposure along the Irish coastline. The map visually represents these differences and guides coastal management and planning efforts in addressing both frequent moderate extremes and rare severe events.

*BM Focus on Annual Extremes*

The BM highlights the most extreme event in each year which might result in a lower return level for frequent events. It may be smoothing out the impact of several extreme events in a year, providing a more conservative estimate for the lower return periods (5-yr and 2-yr). This may imply that regions with infrequent but very severe events appear less vulnerable in the BM analysis, even if they experience multiple significant extreme events per year.

*Practical Implications:*

For regions such as the South cluster, identified as the most vulnerable per the BM method, the focus should be on preparing for rare, sever surges. In contrast, the POT recognises regions which are vulnerable to more frequent and moderate extremes, and so the focus for the North cluster, which has the most exceedances at a 2-yr return level, should be to prepare for moderate but frequent events. This dual approach helps in crafting a more nuanced strategy around coastal management which splits regions into areas which face different types of surges.

8. References

   [List all the sources referenced in the report, following a consistent citation style.]

***Page 1***

EPA Climate Ireland | Coastal Flooding   
Environmental Protection Agency (EPA). Available at: https://www.climateireland.ie/impact-on-ireland/climate-hazards/coastal-flooding/ (Accessed: 19 June 2024).

*Irish Coastal Protection Strategy Study (ICPSS)* (2020) *Government of Ireland*. Available at: https://www.gov.ie/en/publication/eed0fb-irish-coastal-protection-strategy-study-icpss/ (Accessed: 24 July 2024).

Press statement census 2016 results profile 2 - population distribution and movements  
Central Statistics Office (CSO), 2019. Available at: https://www.cso.ie/en/csolatestnews/pressreleases/2017pressreleases/pressstatementcensus2016resultsprofile2-populationdistributionandmovements/ (Accessed: 19 June 2024).

Inland flooding  
Environmental Protection Agency (EPA). Available at: https://www.epa.ie/our-services/monitoring--assessment/climate-change/climate-ireland/impact-of-climate-change-on-ireland/climate-hazards/inland-flooding/ (Accessed: 19 June 2024).

Report of the Inter-Departmental Group on National Coastal Change Management Strategy  
Department of Housing, Local Government and Heritage, 2023. Available at: https://www.gov.ie/en/publication/a43f3-report-of-the-inter-departmental-group-on-national-coastal-change-management-strategy/ (Accessed: 19 June 2024).

Hurricane Ophelia: What to know  
Lui, K., 2017. Time. Available at: https://time.com/4983412/ophelia-atlantic-storm-hurricane-ireland-united-kingdom/ (Accessed: 19 June 2024).

Storm Ophelia caused almost €70m worth of damage  
Towey, N., 2018. The Irish Times. Available at: https://www.irishtimes.com/news/environment/storm-ophelia-caused-almost-70m-worth-of-damage-1.3661365 (Accessed: 19 June 2024).

Ophelia report  
Met Éireann, 2017. The Irish Meteorological Service. Available at: https://www.met.ie/ophelia-report (Accessed: 19 June 2024).

Major weather events  
Met Éireann. The Irish Meteorological Service. Available at: https://www.met.ie/climate/major-weather-events (Accessed: 20 June 2024).

King tides: What explains high water threatening global coasts?  
Woo, M., 2014. National Geographic. Available at: https://www.nationalgeographic.com/science/article/140130-king-tides-high-tides-sea-level-rise-science/ (Accessed: 20 June 2024).

Seiche  
Encyclopædia Britannica. Available at: https://www.britannica.com/science/seiche (Accessed: 20 June 2024).

Climate variability: North Atlantic Oscillation  
Dahlman, R.L.A.L., 2009. NOAA Climate.gov. Available at: https://www.climate.gov/news-features/understanding-climate/climate-variability-north-atlantic-oscillation (Accessed: 20 June 2024).

Computation-efficient parameter estimation for a high-resolution global tide and surge model  
Wang, X., et al., 2021. Journal of Geophysical Research: Oceans, 126(3). Available at: https://doi.org/10.1029/2020jc016917.Stammer, D., Ray, R.D., Andersen, O.B., Arbic, B., Bosch, W., Carrere, L., Cheng, Y., Egbert, G., Fok, H.S., Green, J., Huber, B., Lemoine, F., Lyard, F., Morison, J., Müller, M., Richman, J., Shriver, J.F., Shum, C.K., Taguchi, E., Yi, Y. and Zahel, W., 2014. Accuracy assessment of global barotropic ocean tide models. Reviews of Geophysics, 52(3), pp.243-282.

Oppenheimer, M, et al., 2019: Sea Level Rise and Implications for Low-Lying Islands, Coasts and Communities. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate [H.-O. Pörtner, D.C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, N.M. Weyer (eds.)].

***Page 2***

Dacre HF, Pinto JG (2020) Serial clustering of extratropical cyclones: a review of where, when and why it occurs. NPJ Clim Atmos Sci 3:1–10. <https://doi.org/10.1038/s41612-020-00152-9>

Pickands, J. (1975). "Statistical Inference Using Extreme Order Statistics". The Annals of Statistics, 3(1), 119-131.

9. Appendices

   [Include any additional supplementary information that supports the report but is not necessary in the main body. This should include your code.]

Table X: Estimated parameter covariance matrix for Block Maxima method.

A black screen with white text

Description automatically generated

Table V: Estimated parameter covariance matrix for POT method.

A black background with white text and numbers

Description automatically generated

Table Y: Block maxima parameter estimations with 95% confidence intervals.

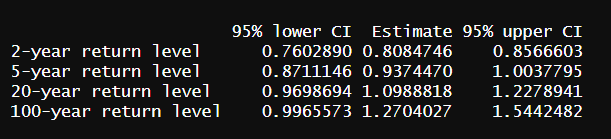


Table W: POT parameter estimations with 95% confidence intervals.

