Estimating Environmental Extremes: New Orleans

Rochelle Marie Firth

April 29, 2009

Abstract

Acknowledgements I would like to thank ..... for supplying the storm surge and wave height data used in this report.

Contents

Chapter 1

Introduction

1.1 Motivation

Early morning, Monday 29 August 2005, Hurricane Katrina made landfall on the Gulf of Mexico shores, bringing devastation to the states of Louisiana, Mississippi and Alabama. Its intense winds, high rainfall, waves and storm surge impacted the greatest on New Orleans and southeast Louisiana. This report considers, in particular, the destruction caused to the city of New Orleans and how it might have been prevented.

New Orleans was built upon low lying marshland along the Mississippi River, with levees and floodwalls consequently being built around the city to protect against flooding. Initially, it was thought that the city had weathered the worst of the storm. However, within hours of the storm passing, it emerged that many of the floodwalls had been overtopped and several key levees had been breached in more than 50 locations [1]. As a result, billions of gallons of water from the Gulf of Mexico and surrounding lakes flowed into New Orleans, leaving an estimated 80 percent of the city under flood water – more than ten feet deep in some parishes [2].

“This catastrophic failure of the city’s hurricane protection system represents one of the nation’s worst disasters ever” [2]. A year later, 1118 people were confirmed dead in Louisiana and another 135 people are still missing. Approximately 400 thousand people evacuated the city as the storm advanced, with over half having not returned; while 50 000–100 000 failed to pay heed to the warnings or were unable to flee [3]. Direct damage to property has been estimated at $21 billion, with another $6.7 billion estimated for public infrastructure damage. Thousands of homes were destroyed, nearly 124 thousand jobs were lost and the region’s economy was crippled. [ref paragraph - 2]

A report by the American Society of Civil Engineers Hurricane Katrina external Review Panel [2] revealed two direct causes for the levee breaches: levees with concrete floodwalls (called I-walls) collapsed because of a design flaw – the margin of safety used in the design process was too low and it did not account for the variability in the strength of soft soils beneath and adjacent to the levees; many levees were overtopped and the water eroded the structures away – despite some overtopping of levees to be expected in a major storm, several of these levees

were constructed of highly erodible soil but were not armoured or protected against erosion [2]. Furthermore, some levees were built 1–2 feet lower than the intended design elevation, despite the acknowledged fact that New Orleans is subsiding (sinking) [2]. Had engineers regularly reviewed the city's subsidence level and accounted for such problems in the design, the hazardous waters may have been held at bay.

Thus, the motivation for this investigation is very real. The design of a levee is made up of several components such as the composition of surrounding land, the strength and depth of its base, the material to use and its height. This report will consider the fatal error resulting from using incorrect data and statistical models and techniques to predict levee elevations, causing many levees to not be built high enough.

1.2 The Data

The data for analysis consists of hourly maximum storm surges (metres) with corresponding maximum wave heights (metres) observed at Shell Beach (SHBL1) in Louisiana. This stations is one of a network of dozens of moored buoys and manned observation stations set up under the Coastal-Marine Automated Network (C-MANN) programme initiated in the 1980s (Figure 1.1). This network collects information on other environmental variables such as atmospheric pressure, wind speed and direction, number of sunshine hours per day and rainfall maxima.

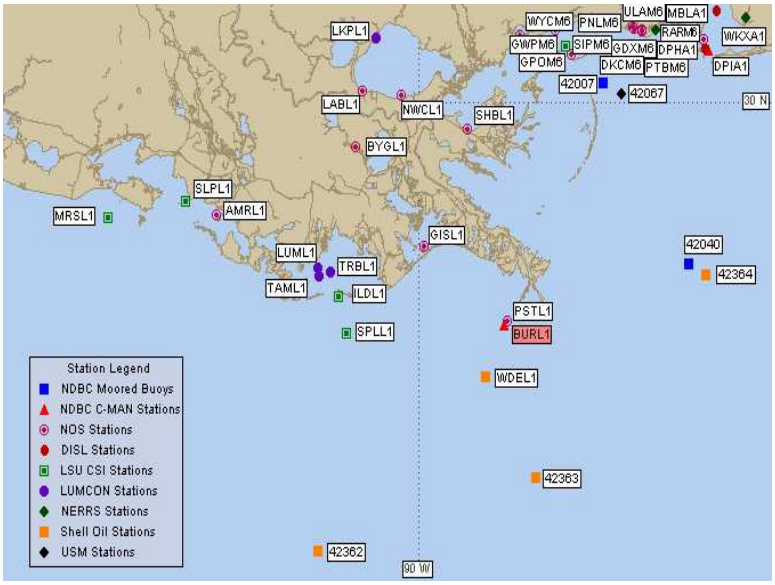

Figure 1.1: Locations of observation stations

For the purposes of this report, the data have been restricted to nine years from 1996 to 2004 (inclusive) so that we can examine whether the recorded extremes that occurred during Hurricane Katrina are captured in our analysis. The hurricane season runs from June through to November and, since this is an extreme value analysis, the non-hurricane seasons have been

discarded. Consequently, the data has been reduced to years consisting of six months worth of data. This gives 183 days per year, meaning there are 4392 hourly observations per year. Thus, in total, there are 39 528 observations of each variable available for analysis.

This report will begin by demonstrating the application of different models to a single data set, in this case sea surges, in order to show the potential short-fallings of models that are currently being used in design procedures and the benefits that might be gained from using alternative models and techniques. Following this, a bivariate analysis will be performed on two datasets – storm surge and wave height.

1.3 What is a Storm Surge?

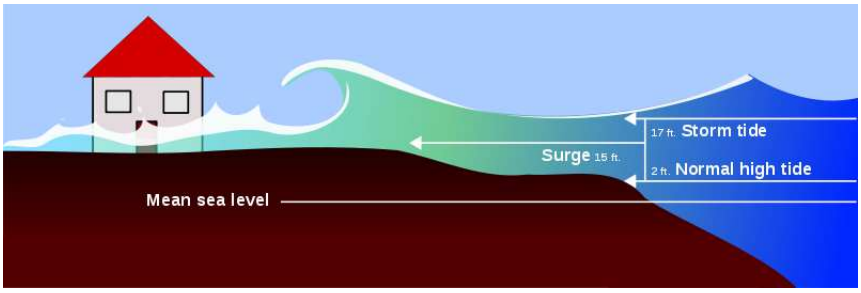

Figure 1.2: Effect of a storm surge

The sea level is made up of four components at any one time: mean sea level, tide, waves and surge. A storm surge is generated by wind (friction) and air pressure which perturbs the water surface, forcing it towards the shore. ”This advancing surge then combines with the normal tides to create the hurricane storm tide, which can increase the mean water level 15 feet or more. In addition, wind-driven waves are superimposed on this storm tide.” [4]. Severe flooding in coastal areas is particularly likely, even more so when the storm tide coincides with the normal high tide.

Since much of New Orleans lies below sea level, the danger of storm tides is extremely threatening. The level of a storm surge is also defined by the slope of the continental shelf. A shallow coastal slope will allow a greater surge to overpower the coastal area. However, even coastlines with steep continental shelves are at threat – “water weighs approximately 1700 pounds per cubic yard and extended pounding by frequent waves can demolish any structure not specifically designed to withstand such forces.” [4].

1.4 Initial Analysis of Data

The New Orleans storm surge dataset will be analysed throughout the report. Figure 1.3 displays a time series plot of the data, a histogram and a plot of the time series against itself at lag 1. From the graphs, it is clear that the time series has successfully had any potential seasonal

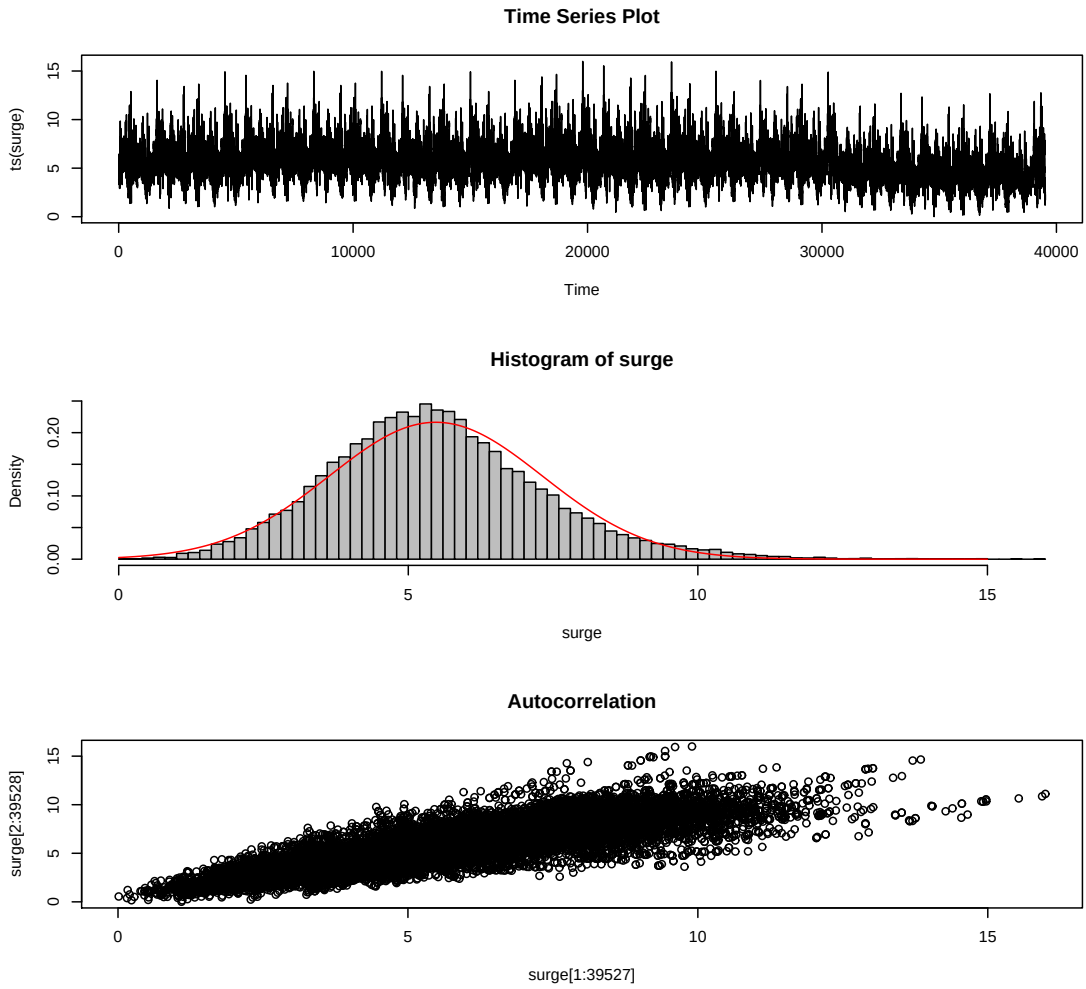

Figure 1.3: Graphical display of surge data

variability removed by the exclusion of non-hurricane months. There appears to be no evidence of trend either. However, there does seem to be strong temporal dependence between successive observations, which will be dealt with later on.

Chapter 2

Extreme Value Theory

2.1 Introduction to Extreme Value Theory

As we have seen, statistical modelling of extreme weather has a practical motivation: reliability. Structures that are built need to have a good chance of surviving the environment for the whole of their working life. Thus, the ability to estimate what the strongest wind or highest tide, for example, will be over some fixed period of future time is essential. The only sensible way of doing this is to use data on the variable of interest and fit a suitable statistical model.

The models used in extreme value theory are motivated by asyptotic theory, since extreme value modelling has a central theoretical result analogous to the Central Limit Theorem. Suppose $X_1, X_2, \ldots$ is an independent, identically distributed sequence of random variables. Define

$$M_n = \max\{X_1, \ldots, X_n\}.$$

We are interested in the limiting distribution of M_n as $n \to \infty$. As with the mean, $\bar{X}$, of $\{X_1, \ldots, X_n\}$, this limiting distribution is degenerate, meaning we need to work with a normalised version.

Extremal Types Theorem
If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$Pr\{(M_n - b_n)/a_n \leq z\} \to G(z) \quad \text{as} \quad n \to \infty,$$

where G is a non-degenerate distribution function, then G belongs to one of the following families:

***I*:** $G(z) = \exp\left\{-\exp\left[-\left(\frac{z-\beta}{\gamma}\right)\right]\right\}, \qquad -\infty < z < \infty$

***II*:** $G(z) = \exp\left\{-\left(\frac{z-\beta}{\gamma}\right)^{-\alpha}\right\}, \qquad z > \beta; \quad [G(z) = 0, z \leq \beta]$

***III*:** $G(z) = \exp\left\{-\left[-\left(\frac{z-\beta}{\gamma}\right)^{\alpha}\right]\right\}, \qquad z < \beta; \quad [G(z) = 1, z \geq \beta]$

for parameters $\gamma > 0$, and $\beta, \alpha > 0$.

Families *I*, *II* and *III* are generally referred to as Gumbel, Fréchet and Weibull, respectively. Fortunately, they can be combined into a single family known as the Generalised Extreme Value distribtion (GEV) which has c.d.f.

$$G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}, \tag{2.1}$$

defined on the set $\{z : 1+\xi(z-\mu)/\sigma > 0\}$; where $\mu, \sigma > 0$ and ξ are the location, scale and shape parameters, respectively. Equation (2.1) correspsonds to families *I*, *II* and *III* in the cases when $\xi = 0$, $\xi > 0$ and $\xi < 0$, respectively.

Now the Extremal Types Theorem can be restated with (2.1) as the limiting form, which provides the basis for the first modelling approach.

2.2 The Block Maxima Approach

The New Orleans storm surge data was divided into blocks of size $n = 30$ days, roughly corresponding to calendar months, each from which the maximum observation was extracted.

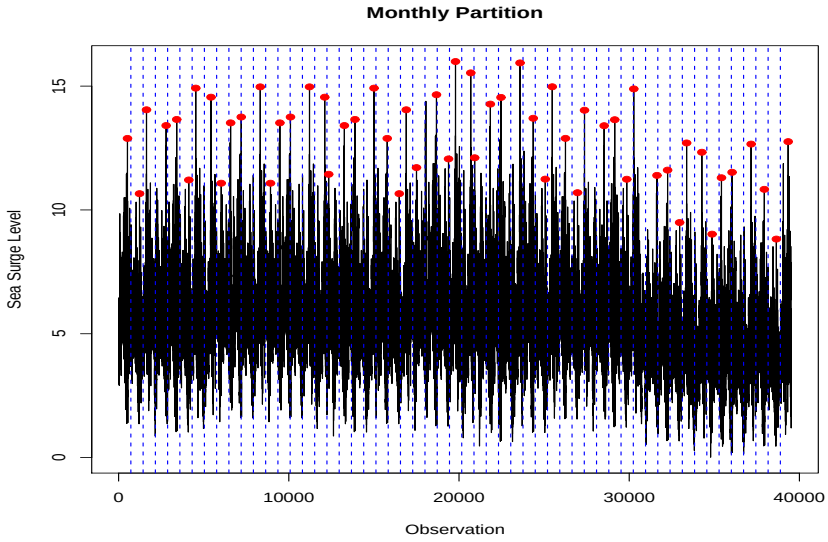

Figure 2.1: Block Maxima Plot of Data

The GEV distribution was then fitted to the resulting sequence of extracted block maxima, $M_{(1)}, \ldots, M_{(N=55)}$, using maximum likelihood estimation, obtaining the following parameter estimates:

$$\hat{\mu} = 12.444\ (0.273) \quad \hat{\sigma} = 1.849\ (0.208) \quad \hat{\xi} = -0.464\ (0.093).$$

Using the standard errors shown in brackets, this gives Wald confindence intervals of $(11.909, 12.979)$, $(1.442, 2.256)$ and $(-0.646, -0.282)$ for the parameters $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}$, respectively.

The goodness-of-fit of the GEV model on the storm surge data was assessed by considering four diagnostic plots, shown in figure 2.2 below.

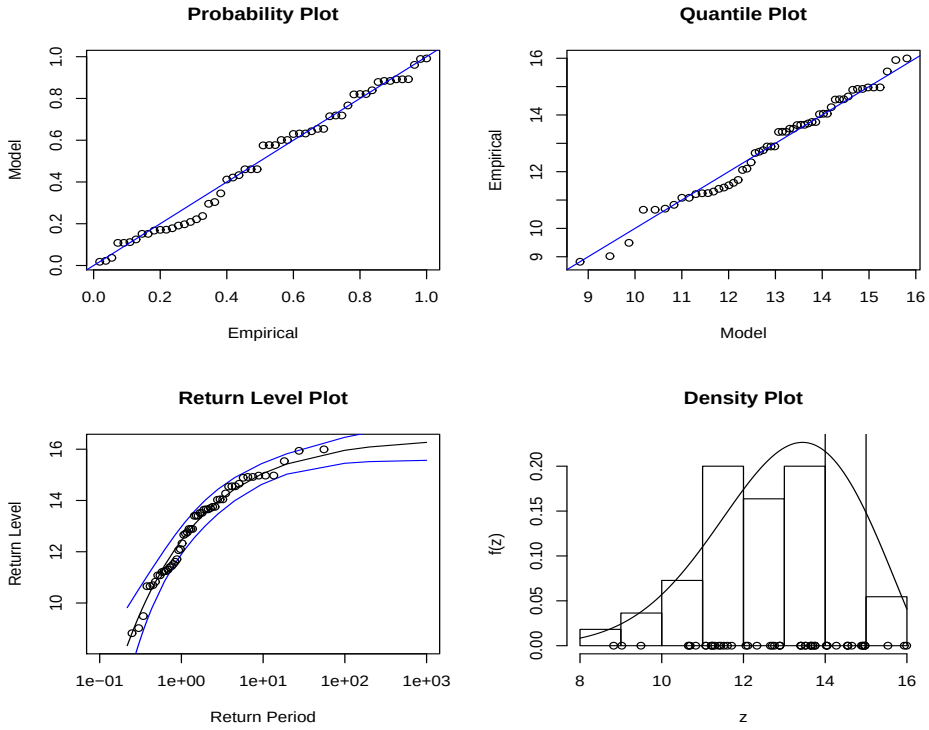

Figure 2.2: Diagnostic plots for GEV fit to storm surge annual maxima

The probability plot shows the fitted value fo the c.d.f. plotted against the empirical values of the c.d.f. for each data point. The quantile plot shows the empirical quantile plotted against the fitted quantile for each data point. This plot contains the same information as the probability plot but expressed on a different scale. In both plots, the points are sufficiently close to linearity to lend support to the GEV model.

The return level plot shows the return level, with error bars, plotted against the return period. Here, each data point defines a sample point. EXPLAIN. Finally, the density plot shows the fitted p.d.f. superimposed on a histogram of the data. EXPLAIN.

The 99^{th} percentile in the distribution of annual maxima is known as the 100 year return level. The fitted value of this is easily obtained on inversion of Model (2.1):

$$q_{100} = 15.957\ (0.467).$$

Usually, a symmetric Wald confidence interval would be constructed using the above standard error. However, this has been proven rather inaccuarate, since it assumes the limiting quadratic

behaviour of the likelihood surface near the maximum. In actual fact, the likelihood surface is very asymmetrical. To account for this asymmetry, the method of profile likelihood is used to calculate confidence intervals.

Profile likelihood...??
A plot of the profile log-likelihood for the storm surge annual maxima is shown below in figure 2.3. The MLE of the 100 year return level has been overlayed, along with the upper 95% Wald confidence interval.

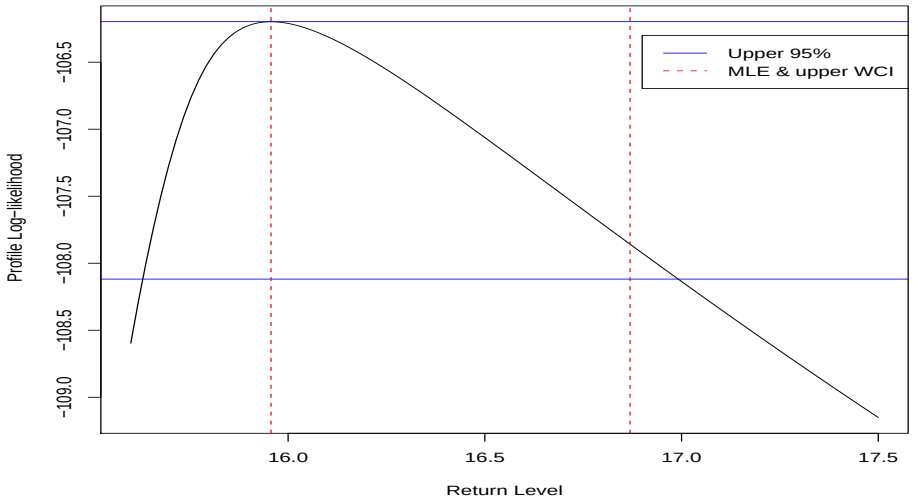

Figure 2.3: Profile log-liklihood for q_{100}

From figure 2.3, the 95% confidence interval is obtained by reading off the points of intersection, leading to a confidence interval for q_{100} of $(15.63, 16.99)$. Compared with the Wald confidence interval of $(15.04, 16.87)$, the profile likelihood interval is slightly smaller in width and is shifted to the right, corresponding to the skewness observed.

2.3 Exceedances over Thresholds Approach

Threshold methods tend to be a more natural way of determing whether an observation is extreme, since all values greater than a specified (high) threshold u are considered, thus, allowing a more efficient use of the data available.

Theorem
If it exists, any limiting distribution as $u \to \infty$ of $(X - u|X > u)$ is of Generealised Pareto Distribution (GPD) form (setting $Y = X - u$):

$$H(y) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)_+^{-\frac{1}{\xi}}, \tag{2.2}$$

where $a_+ = \max(0, a)$ and σ $(\sigma > 0)$ and ξ $(-\infty < \xi < \infty)$ are scale and shape parameters respectively.

Simlar to the GEV, the GPD exists for $\xi = 0$ by taking the limit of (2.2) as $\xi \to 0$, giving:

$$H(y) = 1 - \exp\left(\frac{-y}{\sigma}\right),$$

defined for $y > 0$. This shows that when $\xi = 0$, the GPD is the Exponential distribution with mean equal to the scale parameter σ $(\sigma > 0)$.

The return levels under the threshold excess approach can be calculated based on the following theory. If the GPD is a suitable model for the exceedances of some threshold u, then for $x > u$,

$$Pr\{X > u | X > u\} = \left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]^{-\frac{1}{\xi}}.$$

It follows that

$$Pr\{X > x\} = \lambda_u \left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]^{-\frac{1}{\xi}},$$

where $\lambda_u = Pr\{X > u\}$. So, the level x_m that is exceeded once every m observations is the solution of

$$\lambda_u \left[1 + \xi\left(\frac{x_m - u}{\sigma}\right)\right]^{-\frac{1}{\xi}} = \frac{1}{m}.$$

Rearranging this equation gives

$$x_m = u + \frac{\sigma}{\xi}\left[(m\lambda_u)^{\xi} - 1\right],$$

conditional on m being large enough to ensure that $x_m > u$. Now, if there were n_y observations per year, setting $m = N \times n_y$ would give the N-year return level as

$$z_N = \mu + \frac{\sigma}{\xi}\left[(Nn_y\lambda_u)^{\xi} - 1\right]$$

or, when $\xi = 0$,

$$z_N = \mu + \sigma \log\left(Nn_y\lambda_u\right),$$

with standard errors being obtained using the delta method.

In order to fit the GPD to the storm surge data, a threshold u_0 must be chosen which is high enough so that the GPD will be a good model for $(X - u_0 | X > u_0)$. If the GPD is the correct model for all the exceedances x_i above some high threshold u_0, then the mean excess plotted against $u > u_0$ should give a linear plot. Hence, by producing this plot (known as a mean residual life plot) for values of u starting at zero, a reasonable value for u_0 may be chosen. Figure 2.4 shows the mean residual life plot for the storm surge data, followed by a time series plot of the data with the chosen threshold superimposed.

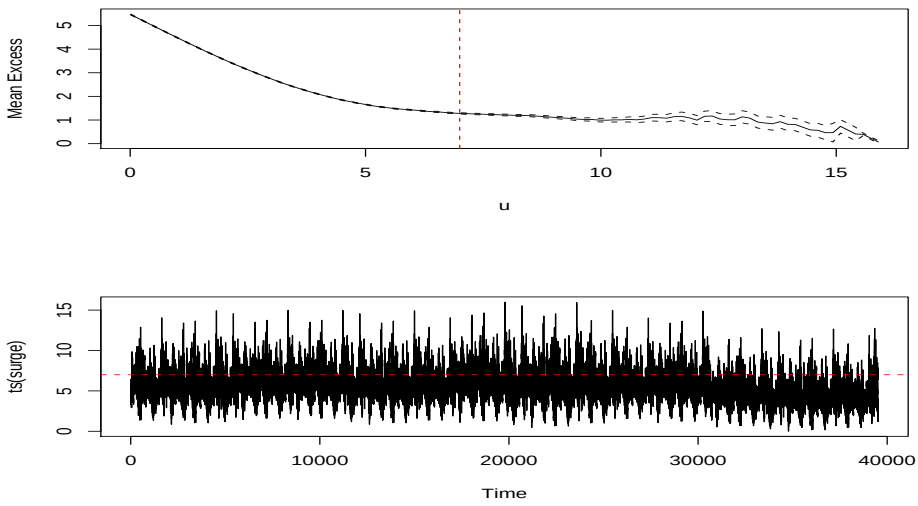

Figure 2.4: Mean residual life plot and threshold superimposed on time series plot of data

From observation of figure 2.4, linearity seems to begin around the threshold value of $u_0 = 7$. Based on this threshold value, model (2.2) turns out to work reasonably well, giving 7343 exceedances $x_i; i = 1, ..., 7343$. Using maximum likelihood methods, the GPD is then fitted to the values $(x_i - u)$, giving

$$\hat{\sigma} = 1.383\ (0.022) \quad \hat{\xi} = -0.079\ (0.010).$$

Using the standard errors shown in brackets, this gives Wald confidence intervals of $(1.340, 1.425)$ and $(-0.099, -0.059)$ for the parameters $\hat{\sigma}$ and $\hat{\xi}$, respectively.

Again, the goodness-of-fit of this model was assessed by considering the same four diagnostic plots as before, shown in figure 2.5.

The probability and quantile plots show the points to be tremendously close to linearity, thus lending significant support to the GPD model for excesses over u. The return level plot shows the return level, with error bars, plotted against the return period. Here, each data point defines a sample point. EXPLAIN. Finally, the density plot shows the fitted p.d.f. superimposed on a histogram of the data. EXPLAIN.

The 100-year return level can be estimated by assuming a uniform rate of exceedances, giving:

$$\hat{q}_{100} = 17.341\ (0.418).$$

Using profile likelihood, shown below in figure 2.6, the 95% confidence interval for q_{100} is approximately $(16.608, 18.26)$. Overlaid are the maximum likelihood estimates of the 100-year return level and it's upper bound. Compared with the 95% Wald confidence interval of $(16.523, 18.160)$, the profile likelihood interval is of a similar width, but shifted to the right to

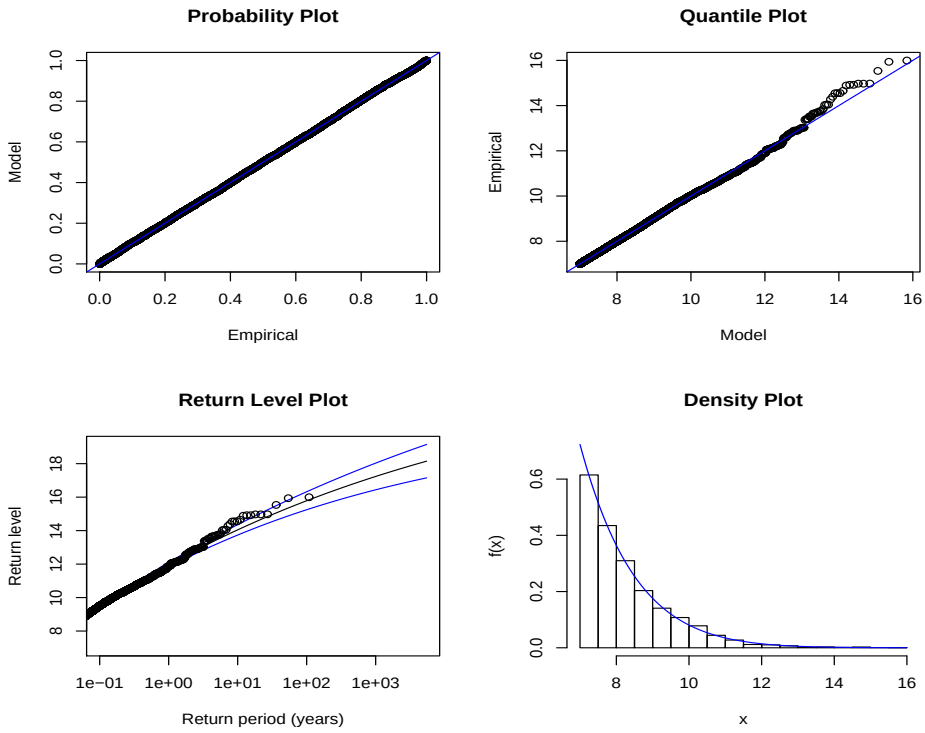

Figure 2.5: Diagnostic plots for GPD model for storm surge data

reflect the skewness obvserved.

In order to determine whether the GPD with shape paramter $\hat{\xi}$ and scale parameter $\hat{\sigma}$ is the correct model for the excesses over $u_0 = 7$, then for any threshold $u > u_0$ the excesses will also be GPD with the same shape and scale parameters:

$$\sigma_u = \sigma_{u_0} + \xi\left(u - u_0\right).$$

It is necessary to use a modified version of the scale parameter,

$$\sigma^* = \sigma_u - \xi u.$$

This now suggests that both σ^* and ξ should be constant over any thresholds greater than u_0 if the excesses $x_i - u$ for $u > u_0$ are modeled using the GPD.

This provides another tool for assessing the initial choice of threshold u_0 and for investigating the stability of the estimates obtained for ξ and σ^*. These parameter stability plots are shown in figure 2.7.

Since the plots for both parameters seem to be reasonably linear, the original choice of $u_0 = 7$ is verified.

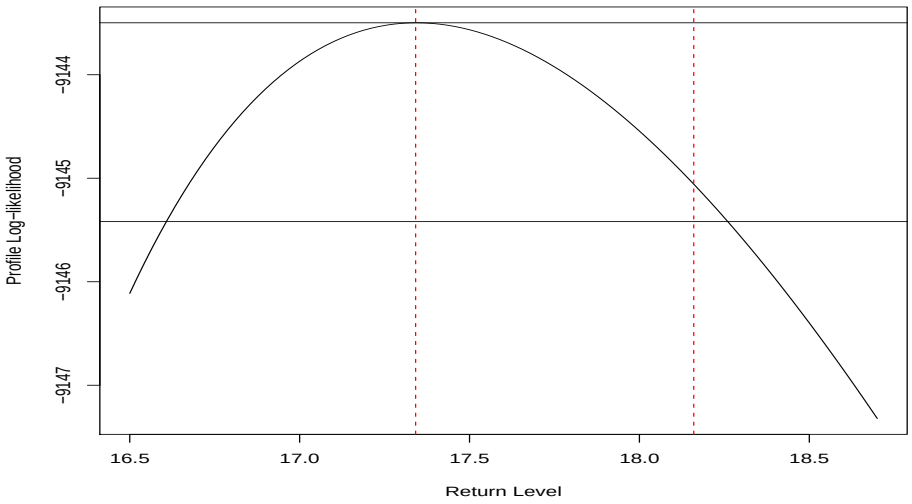

Figure 2.6: Profile log-likelihood for q_{100} based on GPD model

2.4 Modelling Issues

So far, we have considered two approaches to modelling extreme values - the block maxima approach and the threshold exceedances approach. Depsite the robustness of the block maxima approach, it is extremely wasteful of data. Since extreme values are rare in their very nature, it doesn't seem sensible to then exclude all extremes bar one within each block. Especially if the extremes have been collected quite infrequently and, thus, dividing the data into blocks may result in a small sample of maxima from the total data set. Therefore, it would seem fruitful to use the threshold exceedances approach which will allow increased precision due to the inclusion of more extremes and, in turn, providing a larger set of extremes from which inferences can be made.

That's all well and good, but there are several issues that need to be addressed under the threshold approach. Firstly, how should a threshold be selected to identify observations as extreme? The method of observing a mean residual life plot for linearity is rather subjective and leaves room for contradcition from person to person. Although, the choice of threshold can be supported by parameter stability plots. OTHER WAY?

Secondly, the asymptotic results previously described assume the underlying process to be independent and identically distributed (i.i.d.). In practice, there is usually the issue of local temporal dependence, where successive values of a times series are dependent, but values further apart are independent (to a good approximation). Hence, consecutive block maxima will often be far enough apart to be deemed indepent observations, while consecutive exceendances will probably not (depending on how the data was collected). Temporal depedence is an unrealistic assumption, since extreme conditions often persist over several consecutive observations, thus, questioning the suitability of models such as the GEV and GPD. This issue will be dealt

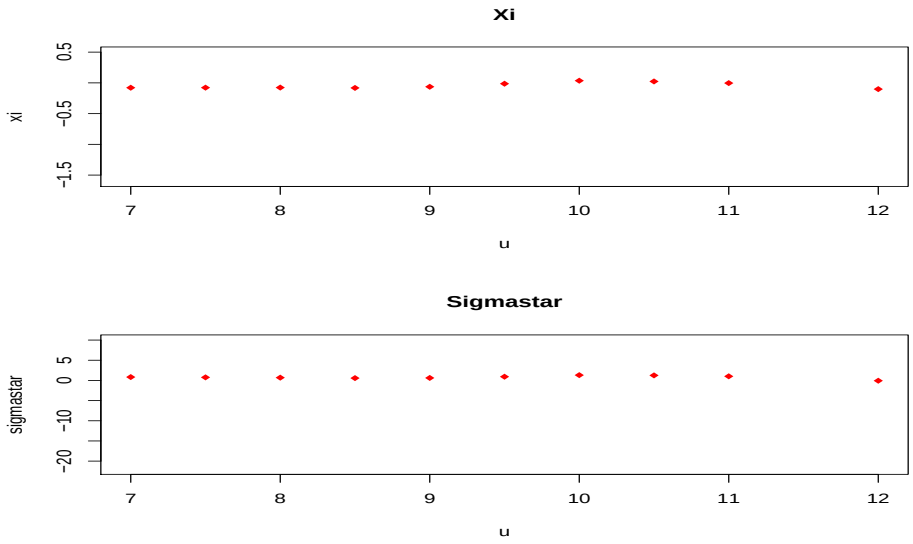

Figure 2.7: Parameter stability plots for GPD for surge data

with in Chapter 3.

There are also issues of long term trends and seasonal variability. As discussed in the initial analysis of the storm surge data set, there appears to be no noticeable trend in the data and seasonal variation has been removed by only considering the acknowledged hurricane months.

?? STATIONARITY

Conclusively, the threshold exceedances approach provides the most beneficial model when considering extreme data. Despite the independence of the set of block maxima extracted using the GEV model, this approach uses the available data inefficiently. The lack of independent data obtained through the threshold exceedance approach can be rectified, and this choice of model is supported further by the large data set often established for inference. Therefore, the remainder of this project will be concerned with the threshold exceedances approach provided by the GPD model.

Chapter 3

Temporal Dependence

3.1 Introduction

As shown in figure 3.1, the storm surge time series has strong autocorrelation, implying substantial temporal dependence.

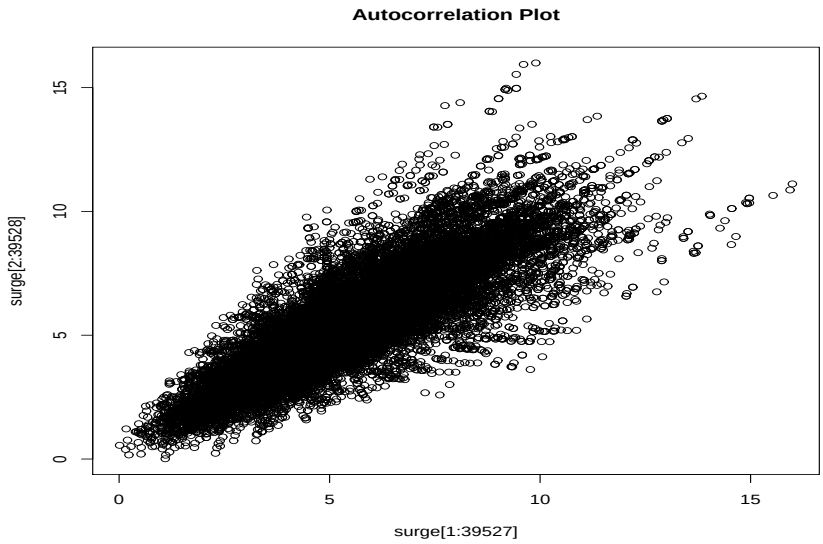

Figure 3.1: Autocorrelation plot of surge data at lag 1

This dependence creates a problem, in that if the GPD is fitted to the extracted set of all threshold exceedances, the likelihoods obtained will be incorrect since they assume the observations to be independent. Fortunately, there are several techniques that have been developed to circumvent this problem. The temporal dependence can be removed by filtering out an (approximately) independent set of threshold exceedances or, having fitted the GPD, the temporal dependence could be ignored and the standard errors adjusted accordingly. Alternatively, the temporal dependence could be explicitly modelled in the process.

While all the approaches are valid, it is the first that is most widely implemented. The third approach is based on multivariate extreme value theory and shall be revisited in more detail in chapter 4. For now, let us consider the options of simply removing the temporal dependence versus ignoring it.

3.2 Comparison of Methods

The most commonly adopted approach employed to remove temporal dependence is a declustering scheme which filters out the set of approximately independent threshold exceedances. One method, which is thought to be the most 'natural', is to identify 'clusters' of extremes using 'runs-declustering'. It is often referred to as the peaks over threshold approach and works in the following manner:

1. Choose an auxiliary 'declustering parameter' κ.
2. A cluster of threshold excesses is terminated as soon as at least κ consecutive observations fall below the threshold u_0.
3. The entire data set should be declusterd in this way.
4. The maximum observation from each cluster is then extracted, forming a cluster peaks data set to which the GPD is fitted.

This approach is relatively easy to implement, however, there are issues surrounding the choice of κ. If κ is too small, the cluster peaks will not be far enough apart to assume independence; whereas, if κ is too large, there will be too few cluster peaks on which to base any inferences made. Furthermore, it has also been shown that parameter estimates can be sensitive to the choice of κ.

Before applying this to the storm surge data, a suitable choice of κ must be found. When considering storm surge data, the presence of wave propagation should be taken into account in an attempt to ensure valid independence between the clusters, since a collection of successive points may all be measurements from the dispersion of a single wave. Therefore, the average propagation of these waves must be taken into consideration when determining κ in order to ensure the separation interval is large enough so that the effects from the propagation of one wave are not captured in consecutive clusters, as the results may still be dependent. Coles and Tawn (1991) suggest that a separation interval of 60 hours should be large enough to safely assume independence between successive clusters and allow for wave propagation. Since variables within the New Orleans data set have been collected hourly, this suggests a separation interval of 2.5 days, i.e. $\kappa = 60$. The threshold $u_0 = 7$ was determined from the mean residual life plot discussed in section 2.2.

Table 3.1 shows the maximum likelihood estimates of the GPD scale and shape parameters σ and ξ, along with the associated 95% Wald confidence intervals, when fitted to the set of cluster peaks using $\kappa = 60$. Shown for comparison, are the corresponding estimates when using all threshold exceedances (ignoring temporal dependence).

	$\hat{\sigma}$	$\hat{\xi}$
Cluster Peaks	6.617	-0.727
95% CI	$(5.228, 8.005)$	$(-0.893, -0.560)$
All Excesses	1.383	-0.079
95% CI	$(1.340, 1.425)$	$(-0.099, -0.059)$

Table 3.1: GPD parameter MLEs and associated confidence intervals

HUGE DIFFERENCE IN PARAMETERS - CLUSTER PEAKS R METHOD WRONG???

Table 3.2 shows the 10, 50 and 100 year return level estimates using both cluster peaks and all exceedances methods with the associated 95% profile likelihood confidence intervals.

	$\hat{q}_{10}$	$\hat{q}_{50}$	$\hat{q}_{100}$
Cluster Peaks	15.824	16.018	16.053
95% CI	$(NA, 16.455)$	$(NA, 16.8)$	$(NA, 16.878)$
All Excesses	15.912	16.938	17.341
95% CI	$(15.405, 16.53)$	$(16.275, 17.765)$	$(16.608, 18.26)$

Table 3.2: 10, 50, 100 year return levels with associated confidence intervals

The above table shows that the estimates for the ten year return period barely differ, but are consistently smaller in the cluster peaks analysis for the 50 and 100 year return periods – substantially so for the 100 year return level. It is with this observation that the importance of using the correct method is amplified. These long-range estimates of return levels are often used as a design requirement in oceangraphic situations, such as the height at which to build sea walls, and are, therefore, extremely influential. If a sea wall, for example, was to be designed to a level specified by an analysis based on the peaks over threshold approach, it could result in considerable under-protection.

Consider the results in table 1. Although there is a slight discrepancy in parameter estimation between the two methods, these discrepancies are negligible. Taking this insignificant difference in estimation, together with the large difference in return period estimation, leads to the conlcusion that surely all exceedances is the more responsible analysis method to be using if the real-life impact such a choice has is considered. As mentioned earlier, however, using this approach means that the standard errors produced for the estimates must be inflated to account for ignoring the short-term temporal dependence. DISCUSS??

3.3 Simulation Study

It is clear that there are significant disagreements in parameter and return period estmations between the cluster peaks method and the all exceedances method. The method that is in practice in most places today is cluster peaks. Although temporal dependence has effectively been removed, this method is still significantly flawed by the under-estimation of return levels – the most crucial result from such an analysis. In order to gain further support for the claim that all

exceedances should be the method of choice, a simulation study was conducted.

The GPD was fitted to a simulated dataset of 10 000 observations, for which the true values of σ (σ^*), ξ and the 1000 and 100 000 return levels (units are a single observation) were known – $\sigma = 1$ ($\sigma^* = 0.301$), $\xi = -0.4$, $q_{1000} = 2.34226$ and $q_{100000} = 2.475$. The strength of the temporal dependence in a dataset is contained within the paramter α, which lies in the range $0 < \alpha < 1$. $\alpha = 1$ represents a lack of temporal dependence, while $\alpha = 0$ signifies complete temporal dependence. For the simulation study, α was set to 0.2, since it is believed that this is representative of real-life environmental data. The threshold for the study was chosen so that 5% of the observations would lie above, giving a 95% threshold value of $u_0 = 1.745278$. The simulation was performed using the peaks over threshold approach, with a separation interval based on $\kappa = 10$, and using the all exceedances method. These baseline values formed the control study, from which several deviations were made.

The simulation was run 1000 times to obtain sampling distributions for the parameters and return periods, along with the standard errors. Alterations to the control study included varying the strength of the temporal dependence α, altering the separation interval of the cluster peaks κ and varying the length of the simulated dataset. Figure 3.2...

Figure 3.3 shows a collection of graphs showing the most significant results, along with tabulated values.

	$\hat{\sigma}$	$\hat{\xi}$	$\hat{q}_{1000}$	$\hat{q}_{100000}$
(True Value)	(0.301)	(-0.4)	(2.34226)	(2.475)
Cluster Peaks	0.468	-0.669	2.289	2.441
95% CI	$(0.370, 0.586)$	$(-0.886, -0.503)$	$(2.231, 2.339)$	$(2.371, 2.494)$
All Excesses	0.305	-0.417	2.332	2.458
95% CI	$(0.258, 0.360)$	$(-0.510, -0.326)$	$(2.274, 2.383)$	$(2.382, 2.524)$

Table 3.3: Values associated with Figure 3.3

STANDARD ERRORS

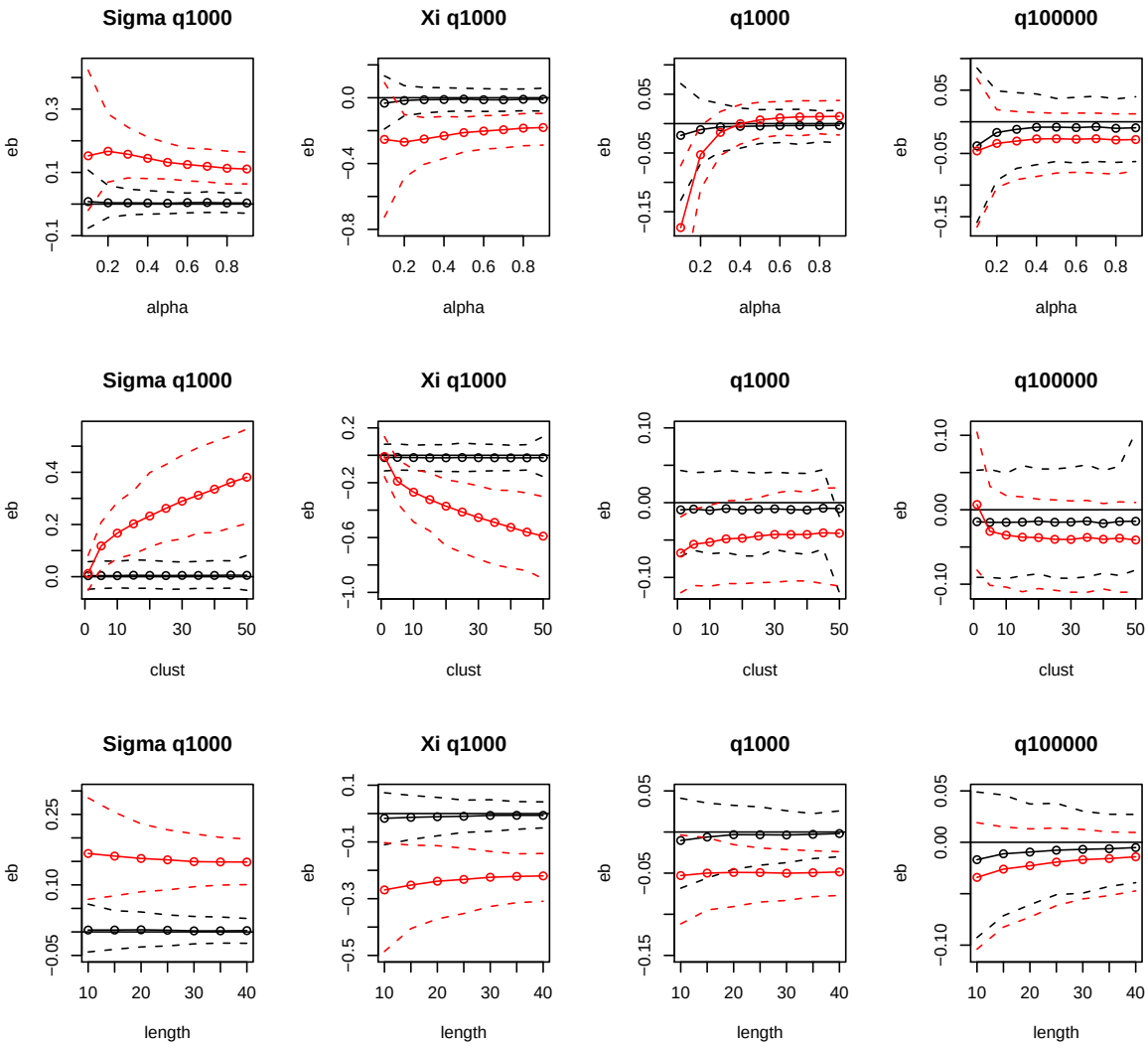

Figure 3.2: Simulation study results

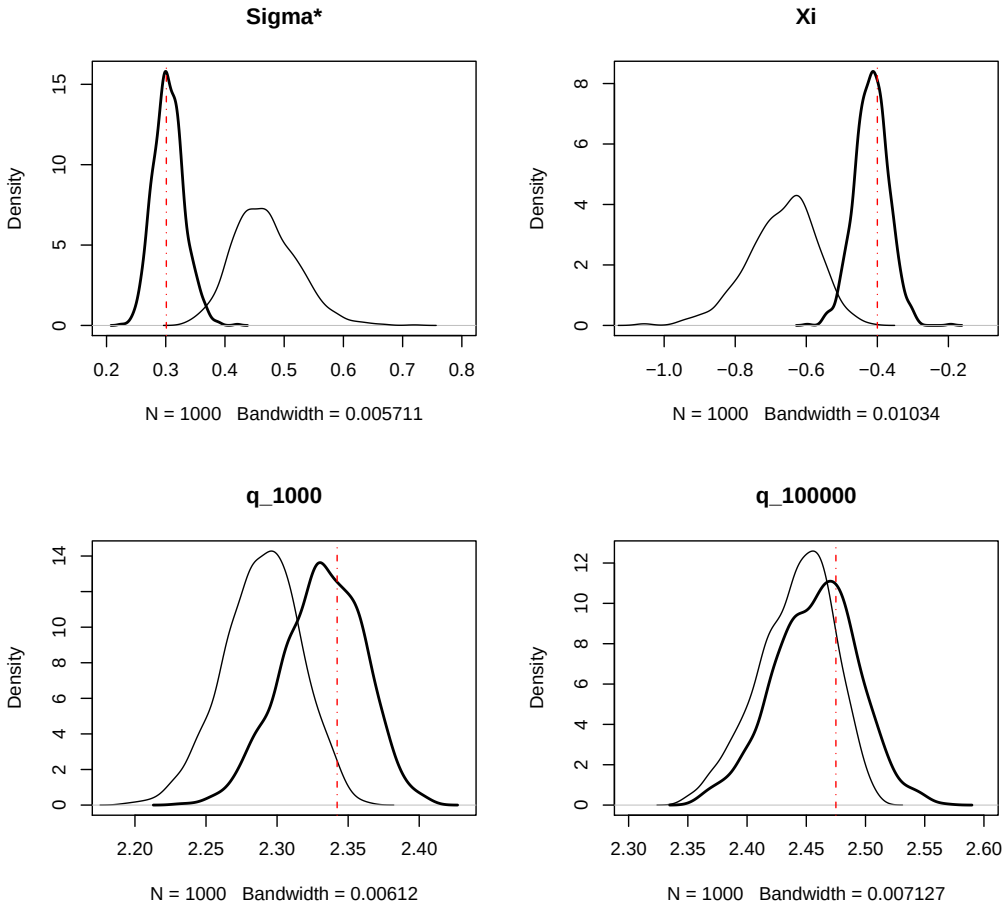

Figure 3.3: Density plots for Control study

Chapter 4

Bivariate Extremes

4.1 Multivariate Extreme Value Theory

Multivariate analysis of variables thought to be dependent on one another is an important part of the inference procedure. In this case, existing wave heights would most certainly have a huge effect on the height of storm surges. Even the ability to compare the storm surges across a range of shoreline locations would be useful. It is, therefore, necessary to consider some multivariate extreme value theory through simplification to the bivariate case, for ease of explanation.

Let $(X_1, Y_1), (X_2, Y_2), \ldots$ be i.i.d. vectors with the distribution function $F(x, y)$. To begin with, consider the componentwise block-maxima

$$M_{x,n} = \max_{i=1,\ldots,n} \{X_i\} \quad \text{and} \quad M_{y,n} = \max_{i=1,\ldots,n} \{Y_i\}.$$

The vector of componentwise maxima is then defined to be $\boldsymbol{M_n} = (M_{x,n}, M_{y,n})$ and we are interested in the limiting behaviour of $\boldsymbol{M_n}$ as $n \to \infty$. It should be noted that $\boldsymbol{M_n}$ will not necessarily be one of the original observations (X_i, Y_i) and that standard univariate extreme value results apply in each margin.

It is conveniently assumed that the X_i and Y_i variables have known marginal distributions – conveniently, GEV(0,1,1) distributions. This distribution is also known as the unit Fréchet distribution, with c.d.f.

$$F(z) = \exp(-1/z), \quad z > 0.$$

This gives rise to a very simple normalisation of maxima:

$$Pr(X_i < x) = Pr\left(\frac{M_{x,n}}{n} < x\right) = \exp\left(-\frac{1}{x}\right), \quad x > 0,$$

with a similar result for Y_i. Now consider the re-scaled vector

$$\mathbf{M_n^*} = \left(\frac{\max_{i=1,\ldots,n}\{X_i\}}{n}, \frac{\max_{i=1,\ldots,n}\{Y_i\}}{n}\right),$$

the margins are unit Fréchet for all n.

Theorem
Let $M_n^* = (M_{x,n}^*, M_{y,n}^*)$ be the normalised maxima as above, where (X_i, Y_i) are i.i.d. with standard Fréchet marginal distributions. If

$$Pr(M_{x,n}^*, M_{y,n}^*) \to G(x,y),$$

then G has the form

$$G(x,y) = \exp\{-V(x,y)\}; \quad x,y > 0$$

where

$$V(x,y) = 2\int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right)\, dH(\omega)$$

and H is a distribtion function on [0,1] satisfying the mean constraint:

$$\int_0^1 \omega\, dH(\omega) = 0.5.$$

Since the GEV provides the complete class of marginal limit distributions, the complete class of bivariate extreme value distributions is obtained as follows. Suppose X and Y are GEV with parameters (μ_x, σ_x, ξ_x) and (μ_y, σ_y, ξ_y) respectively, then unit Fréchet margins are obtained by the transformations

$$\tilde{x} = \left[1 + \xi_x\left(\frac{x-\mu_x}{\sigma_x}\right)\right]^{\frac{1}{\xi_x}} \quad \text{and} \quad \tilde{y} = \left[1 + \xi_y\left(\frac{y-\mu_y}{\sigma_y}\right)\right]^{\frac{1}{\xi_y}}.$$

Hence,

$$G(x,y) = \exp\{-V(\tilde{x},\tilde{y})\}$$

is a bivariate extreme value distribution with the appropriate margins for valid $V(.)$, provided that $\{[1 = \xi_x(x-\mu_x)/\sigma_x] > 0\}$ and $\{[1 + \xi_y(x-\mu_y)/\sigma_y] > 0\}$.

Adapting this theory to the threshold approach for a bivariate observation (X, Y), the aim is now to define bivariate extremes in those observations which exceed a threshold in the respective margins. Focusing on X, the distribution function for the exceedances of a threshold u by a variable X has previously been shown to be:

$$G(x) = 1 - \lambda\left\{1 + \frac{\xi(x-u)}{\sigma}\right\}^{-\frac{1}{\xi}}$$

defined on $\{x - u : x - u > 0\}$ and $\{(1 + \xi(x-u)/\sigma) > 0\}$, where $\xi \neq 0$, $\sigma > 0$ and $\lambda = Pr(X > u)$. This produces a unit Fréchet margin with the transformation:

$$\tilde{X} = -\left(\log\left\{1 - \lambda_x\left[1 + \frac{\xi_x(X-u_x)}{\sigma_x}\right]^{-\frac{1}{\xi_x}}\right\}\right)^{-1}.$$

Applying the analogous transformation to the Y margin gives:

$$\tilde{F}(\tilde{x}, \tilde{y}) = \exp\{-V(\tilde{x}, \tilde{y})\}; \qquad x > u_x, \quad y > u_y,$$

where:

$$V(x, y) = 2 \int_0^1 \max\left(\frac{\omega}{x}, \frac{1-\omega}{y}\right) \, dH(\omega)$$

and H is a distribution function on $[0, 1]$ satisfying the mean constraint:

$$\int_0^1 \omega \, dH\omega = 0.5.$$

4.2 Practical Application

Considering the storm surge and wave height data in a bivariate situation, we can examine the dependence between the two variables. A bivariate plot of the storm surge and wave height data is shown in Figure 4.1 with the corresponding thresholds (obtained from mean residual life plots). There is a strong tendancy for large values of one variable to correspond to large values of the other.

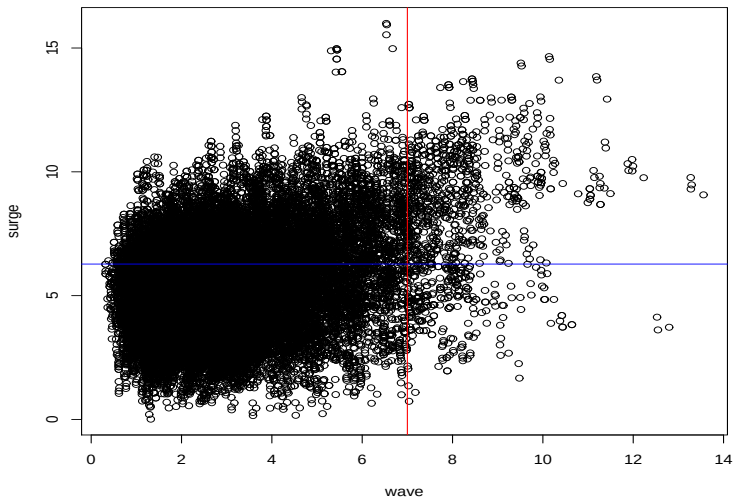

Figure 4.1: Threshold classification of Bivariate Data

Inference for bivariate thresholds is complicated by the possibility that a bivariate pair (x, y) may be an 'exceedance' and yet only exceed the specified thresholds in one of the two components. This can clearly be seen by dividing the plot into four regions:

Region 1: $R_{1,1} = [u_x, \infty) \times [u_y, \infty)$

Region 2: $R_{1,0} = [u_x, \infty) \times (-\infty, u_y)$

Region 3: $R_{0,1} = [-\infty, u_x) \times [u_y, \infty)$

Region 4: $R_{0,0} = (-\infty, u_x) \times (-\infty, u_y)$

This has an effect on the likelihood calculations. The bivariate model described only applies to points located in Region 1, with the density of $\tilde{F}(\tilde{x}, \tilde{y})$ giving the appropriate likelihood component. For the remaining regions, the likelihood component for the points must be censored. ??INCLUDE??

4.2.1 Modelling the Dependence

Bivariate extreme value models contain many families of distributions which can be used to model the dependence structure in the data. The chosen dependence structure has to statisfy the conditions on $H\omega$. There are several possible choices, but this report will focus on only two: the Logisitic model (symmetric) and the Bilogisitic model (assymetric). Both these models are commonly used, yet contrast greatly.

The Logisitic Model
This model has the form

$$G(x, y) = \exp\left\{-\left(x^{-\frac{1}{\alpha}} + y^{-\frac{1}{\alpha}}\right)^{\alpha}\right\},$$

where $x, y > 0$ and $\alpha \in (0, 1]$. As $\alpha \to 1$, the variables are considered to be independent. As $\alpha \to 0$, the variables are considered to be perfectly dependent. Since this model is symmetric, the variables are exchangeable.

The Bilogisitic Model
This model has the form

$$G(x, y) = \exp\left\{x\gamma^{1-\alpha} + y(1-\gamma)^{1-\beta}\right\},$$

where $0 < \alpha, \beta < 1$ and $\gamma = \gamma(x, y; \alpha, \beta)$ is the solution of:

$$(1-\alpha)x(1-\gamma)^{\beta} = (1-\beta)y\gamma^{\alpha}.$$

In this model, independence is achieved when $\alpha = \beta \to 1$ or when either α or β is fixed and the other approaches 1. The extent of asymmetry in the dependence structure is determined by the value of $\alpha - \beta$. If $\alpha = \beta$, this model reduces to the Logisitic model.

Fitting these two models to the wave-surge data obtained the following results tabulated in Table 4.1, with standard errors shown in brackets. These results suggest a fairly weak, yet significant, dependence. A formal likelihood ratio test can be used to compare these models, since the logisitc model is a subset of the bilogisitc model. This lead to a deviance test statistic of 27.12, which is significantly large when compared to a χ^2_1 distribution. Hence, there is significant evidence of asymmetry in the dependence structure, suggesting the bilogisitic model is the more

Model	Log-lik.	$\hat{\boldsymbol{\alpha}}$	$\hat{\boldsymbol{\beta}}$
Logisitic	-38518.02	$0.760(0.006)$	
Bilogisitc	-38504.46	$0.796(0.008)$	$0.688(0.017)$

Table 4.1:

appropriate.

Figure 4.2 confirms the asymmetry of the dependence structure, by displaying the spectral densities of the logisitic and bilogistic models, at the respective values of $\hat{\alpha}$.

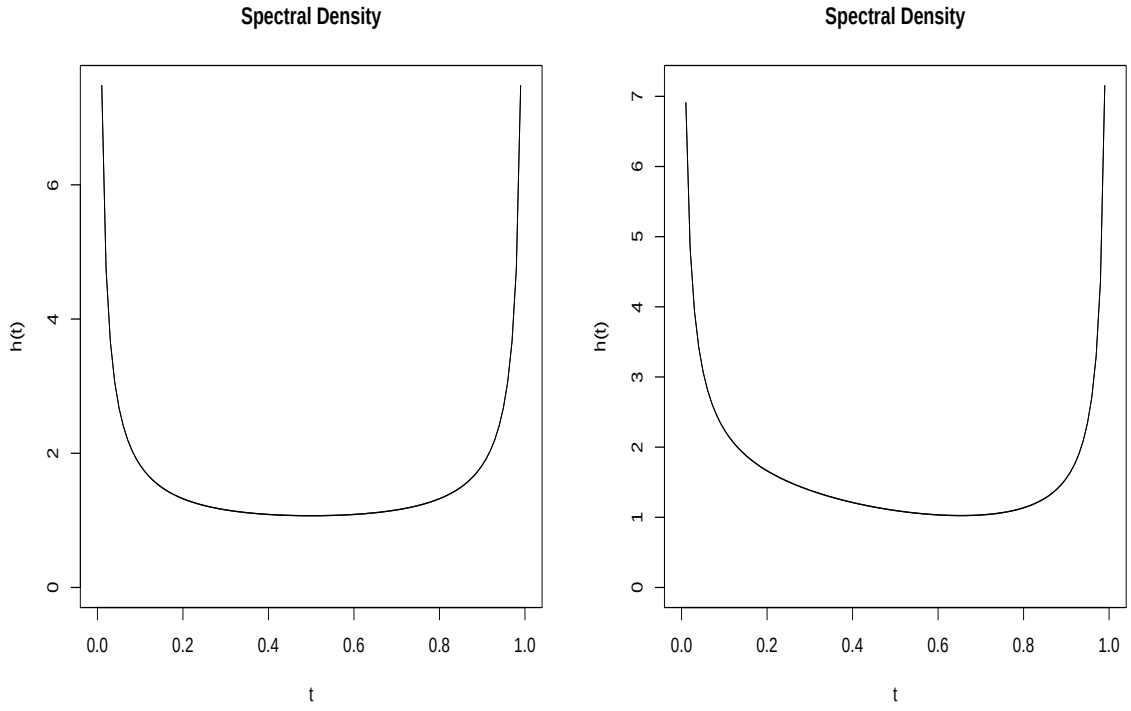

Figure 4.2: Logistic and Bilogisitc model dependence functions, respectively

4.2.2 Examining Asymptotic Dependence

There is a problem when modelling bivariate extremes using the threshold excess model – it is likely to overestimate dependence on extrapolation. The strength of extremal dependence when using the bilogisitic model can be summarised by examining the behaviour of $\chi(u)$, $(0 \leq \chi \leq 1)$. For asymptotically independent variables, $\chi = 0$. The value of χ increases with the strength of dependence at extreme levels, within asymptotically depenedent variables.

There is a corresponding measure, $\bar{\chi}(u)$ $(-1 \leq \bar{\chi} \leq 1)$, which provides a measure of discrimination for asymptotically independent distributions. For asymptotically dependent variables,

$\bar{\chi} = 1$, while for independent variables, $\bar{\chi} = 0$. The value of $\bar{\chi}$ increases with strength of dependence at extreme levels for asymptotically independent variables.

Taken together, $(\chi(u), \bar{\chi}(u))$ provide a summary of extremal dependence. If $\bar{\chi} < 1$, then $\chi = 0$, suggesting that the variables are asymptotically independent. In this case, the value of $\bar{\chi}$ is a more appropriate measure of the strength of extremal dependence. Figure 4.3 displays the empirical estimates of $\chi(u)$ and $\bar{\chi}(u)$ for the wave-surge data.

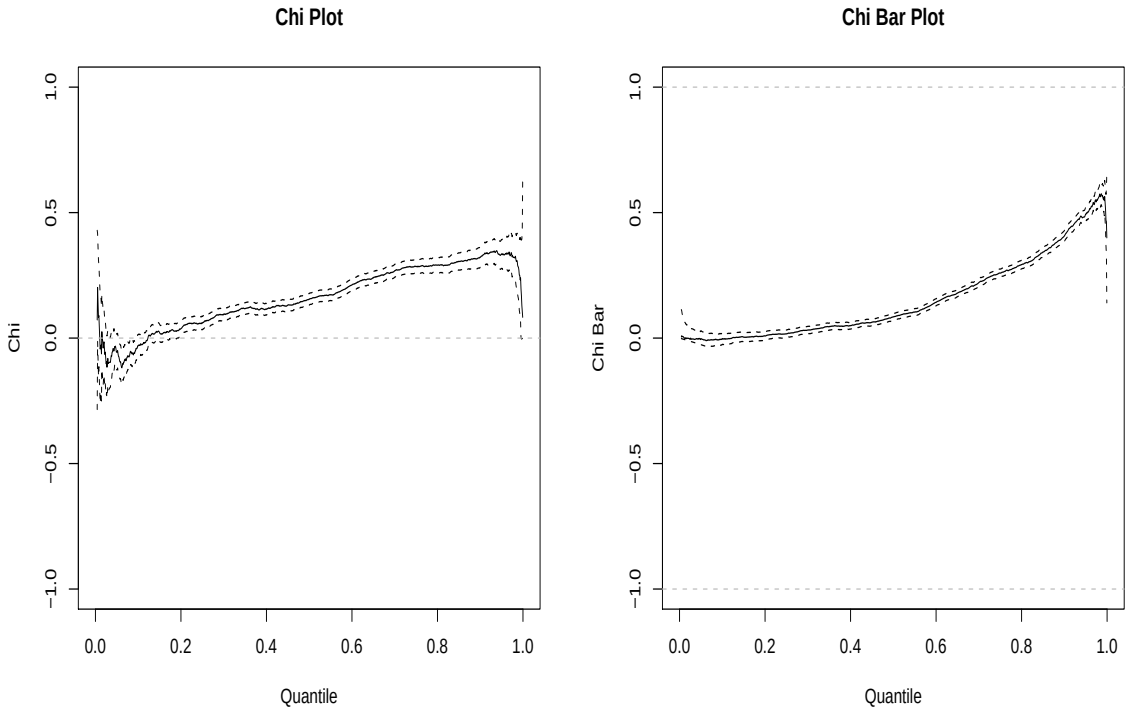

Figure 4.3: Chi and Chi bar plots with 95% confidence intervals

The second graph seems to be consistent with the possibility thet $\bar{\chi} \to 1$ as $u \to 1$. In the first graph, $\chi(u)$ seems to converge to a value of around 0.7. These observations lend support to the use of the bilogisitic model for asymptotic dependence above the threshold $u_0 = 7$.

Chapter 5

Bayesian Inference for Extremes

5.1 Motivation

Throughout this report, parameter estimation has been performed using the general method of maximum likelihood. This method adopts the model with greatest likelihood, such that, of all the possible models, it is the one that assigns the highest probability to the observed data. An alternative and more desirable way to draw inferences from the likelihood function is to use Bayesian techniques.

The appeal of performing a Bayesian anaylsis rather than maximum likelihood is threefold. Due to the scarcity of data, a Bayesian analysis of extreme value data can include other sources of information through a prior distribution. It also outputs a posterior distribution which provides a more complete inference than maximum likelihood analyisis. Furthermore, since the aim of an extreme value anaylsis is to estimate the probabliity of future events reaching extreme levels, predicitive distributions reflect this aleatory uncertainty in future observations and the epistemic uncertainty in the parameters, with all remaining uncertainty properly reflected. The prior predictive distribution then describes our beliefs before we have seen data and the posterior predicitve distribution dsecribes our beliefs afterwards.

5.2 General Theory

The fundamental defining characteristic of Bayesian inference is that parameters are treated as random variables. This is quite different from the frequentist inference, when the data are regarded as coming from a distriution described by parameters of fixed but unknown value.

In general, we can use Bayes' rule to change a prior probability distribution , which expresses our beliefs about parameters before we see the data, to a posterior probability distribution, representing beliefs about the parameters given the data. Suppose we have a prior probability density function for a vector θ of parameters, $f_\theta(\theta)$. Suppose the p.d.f. for a vector $\boldsymbol{Y}$ of observations given $\boldsymbol{\theta}$ is $f_{Y|\theta}(\boldsymbol{y}, \boldsymbol{\theta})$. This latter p.d.f. is treated as a function of $\boldsymbol{\theta}$ once $\boldsymbol{y}$ is observed and is called the liklihood.

Then the posterior p.d.f. is given by

$$f_{\theta|y}(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f_\theta(\boldsymbol{\theta}) f_{Y|\theta}(\boldsymbol{y}|\boldsymbol{\theta})}{\int_\theta f_\theta(\boldsymbol{\theta}) f_{Y|\theta}(\boldsymbol{y}|\boldsymbol{\theta})\, d\theta}.$$

Often it is sufficient to write

$$f_{\theta|y}(\boldsymbol{\theta}|\boldsymbol{y}) \propto f_\theta(\boldsymbol{\theta}) f_{Y|\theta}(\boldsymbol{y}|\boldsymbol{\theta}).$$

That is

$$\textbf{posterior} \propto \textbf{prior} \times \textbf{likelihood}.$$

5.2.1 Predicitive Distributions

Suppose that our beliefs about a parameter θ are represented by the distribution with density function $f_\theta(\theta)$. The distribution of a future observation Y given θ has density function $f_{Y|\theta}(y|\theta)$. The joint density of θ and Y is, therefore,

$$f_\theta(\theta) f_{Y|\theta}(y|\theta).$$

To find the marginal density of Y, we simply integrate out θ:

$$f_Y(y) = \int_\infty^\infty f_\theta(\theta) f_{Y|\theta}(y|\theta)\, d\theta.$$

This marginal distribution is called a predictive distribution of Y. Its mean is the predictive mean and so on.

Most likely, we are interested in the case where we observe some data and then make predicitions about future observations which we have not yet seen. In this case, before we ahve seen the data, we can evaluate the prior predictive distribution:

$$f_Y^{(0)}(y) = \int_{-\infty}^\infty f_\theta^{(0)}(\theta) f_{Y|\theta}(y|\theta)\, d\theta,$$

where $f_\theta^{(0)}(\theta)$ is the prior density of θ. After we have seen data D, we can evaluate the posterior predictive distribution:

$$f_Y^{(1)}(y|D) = \int_{-\infty}^\infty f_{\theta|D}^{(1)}(\theta|D) f_{Y|\theta}(y|\theta)\, d\theta,$$

where $f_{\theta|D}^{(1)}(\theta|D)$ is the posterior density of θ given the data D.

5.2.2 MCMC

Bayesian analysis, therefore, provides a distribution for parameters. Often these distributions are very complicated analytically, so we use stochastic simulation to simulate samples from

the distribution in question. The main idea is to construct a chain such that the stationary distribution is the one we are interested in. We then generate samples from the distribution by simulation the Markov chain. This is known as Markov chain Monte Carlo (MCMC). There are a variety of different techniques, however, we shall use the two most fundamental: the Gibbs sampler and the Metropolis-Hastings algorithm. We will use a hybrid method which uses different Metropolis-Hastings chains to sample from each full conditional at each step of a Gibbs sampling algorithm.

The Gibbs Sampler
This is a way to sample from multivariate distributions based on the ability to simulate from conditional distributions. Suppose the density of interest is $\pi(\theta)$, where $\theta = (\theta_1, \ldots, \theta_d)'$ and that the full conditionals are available for simulation. We denote the conditionals by

$$\pi(\theta_i|\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_d) = \pi(\theta_i, \theta_{-i}), \quad i = 1, \ldots, d.$$

The Gibbs sampler is defined by the following algorithm:

1. Initialise the iteration counter to $j = 1$.
   Initialise the state of the chain to $\theta(0) = (\theta_1^{(0)}, \ldots, \theta_d^{(0)})'$.
   This value must lie in the support of π.

2. Obtain a new value $\theta^{(j)}$ from $\theta^{(j-1)}$ by successive generation of values:

$$\begin{aligned}
\theta_1^{(j)} &\sim \pi(\theta_1|\theta_2^{(j-1)}, \ldots, \theta_d^{(j-1)}) \\
\theta_2^{(j)} &\sim \pi(\theta_2|\theta_1^{(j)}, \theta_3^{(j-1)}, \ldots, \theta_d^{(j-1)}) \\
&\vdots \\
\theta_{d-1}^{(j)} &\sim \pi(\theta_{d-1}|\theta_1^{(j)}, \ldots, \theta_{d-2}^{(j)}, \theta_d^{(j-1)}) \\
\theta_d^{(j)} &\sim \pi(\theta_d|\theta_d^{(j)}, \ldots, \theta_d^{(j)})
\end{aligned}$$

3. Change counter j to $j + 1$ and return to step 2.

This defines a homogeneous Markov chain, since each simulated value depends only on the previous simulated value and not on any previous value or the iteration counter j.

Metropolis-Hastings Sampling
Suppose that $\pi(\theta)$ is the density of interest. Suppose we have some (almost arbitrary) transition kernel $q(\theta, \phi)$. This is used to propose a new realisation ϕ given the current realisation θ. The Metropolis-Hastings algorithm is as follows:

1. Initialise the iteration counter to $j = 1$ and initialise the chain to $\theta^{(0)}$, chosen from somewhere in the support of $\pi(\theta)$.

2. Generate a proposed value ϕ using the transition kernel $q(\theta^{(0)}, \phi)$.

3. Evaluate the acceptance probability $\alpha(\theta^j, \phi)$ of the proposed move, defined by

$$\alpha(\theta, \phi) = \min\left\{1, \frac{\pi(\phi)q(\phi, \theta)}{\pi(\theta)q(\theta, \phi)}\right\}.$$

4. Put $\theta^{(j+1)} = \phi$ with probability $\alpha(\theta^{(j)}, \phi)$; otherwise put $\theta^{(j+1)} = \theta^{(j)}$.

5. Put j to $j + 1$ and go to step 2.

At each stage, a new value is generated from the proposal distribution. This is eacher accepted, in which case the chain moves, or rejected, in which case the chain stays at the same point. Whether the move is accepted or rejected depends on the aceptance probability, which itself depends on the relationship between the density of interest and the proposal distribution.

5.3 Practical Application

Using the GPD as a model for all exceedances over the threshold $u = 7$,

$$X_i \sim GPD(\sigma, \xi) \quad i = 1, \ldots, 7484.$$

When using MCMC methods, the GPD scale parameter is usually re-paramererised by $\eta = \log(\sigma)$ in order to retain the positivity. In the absence of expert prior knowledge regarding the GPD parameters, a 'naive' approach is apdopted, using largely non-informative, independent priors. In this case, the following distributions were used:

$$\begin{aligned} \pi(\eta) &\sim N(0, 10000) \\ \pi(\xi) &\sim N(0, 100), \end{aligned}$$

the large variances impose near-flat priors.

Initialising the MCMC scheme at several different points, the values for both cluster peaks and all excesses are shown in Figure 5.1. The value for cluster peaks were genereated by 50 000 iterations. The burn-in period seems to take about 10 000 iterations, after which the chains appear to converge to the stationary distribution. Deleting these first 10 000 simulations, the remaining 40 000 simulated values can now be treated as dependent realisations, whose marginal distribution is the posterior. The values for all excesses were generated by 10 000 iterations. The burn-in period seems to take about 400 iterations, after which the chains appear to converge to the stationary distribution. Deleting these first 400 simulations, the remaining 9 600 simulated values can now be treated as dependent realisations, whose marginal distribution is the posterior.

Figure 5.2 displays the sampling distributions for the GPD parameters and the 100-year return level, for cluster peaks and all excesses. The corresponding maximum likelihood estimations (red dotted line) have been superimposed on each plot for ease of visual comparison.

Table 5.1 compares the values obtained by the two different analysis methods with 95% credible intervals, the Bayesian estimates have been acheived by taking the posterior medians.

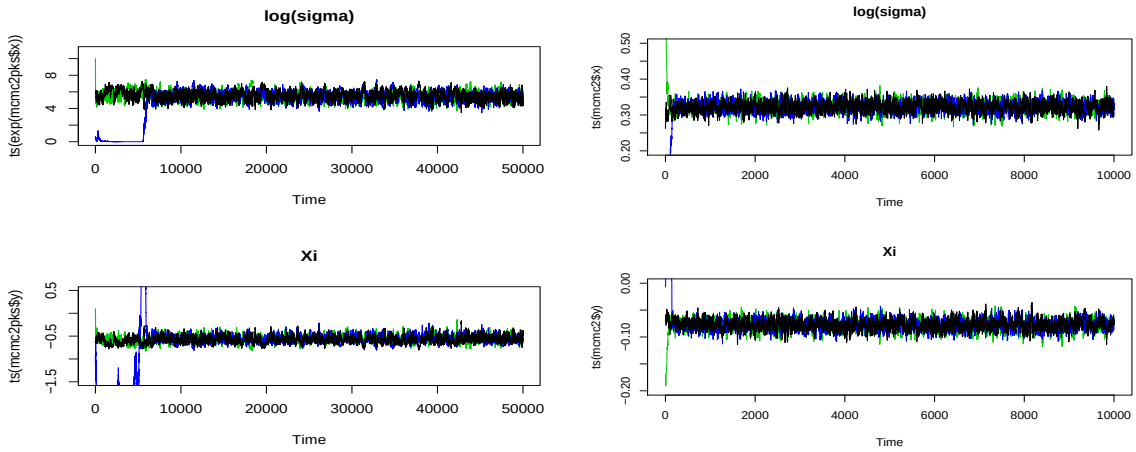

Figure 5.1: MCMC realisations of GPD parameters for clusterpeaks and all excesses, respectively

		$\hat{\sigma}$ (95% CI)	$\hat{\xi}$ (95% CI)	$\hat{q}_{100}$ (95% CI)
Max. Likelihood	**Cluster Peaks**			
	All Excesses	1.383 (1.340,1.425)	-0.079 $(-0.099, -0.059)$	17.341 (16.608,18.26)
Bayesian	**Cluster Peaks**	5.562 (4.543,6.671)	-0.587 $(-0.722, -0.428)$	16.232 (15.963,17.651)
	All Excesses	1.381 (1.339,1.422)	-0.078 $(-0.097, -0.056)$	17.380 (16.642,18.323)
	Predictive			

Table 5.1:

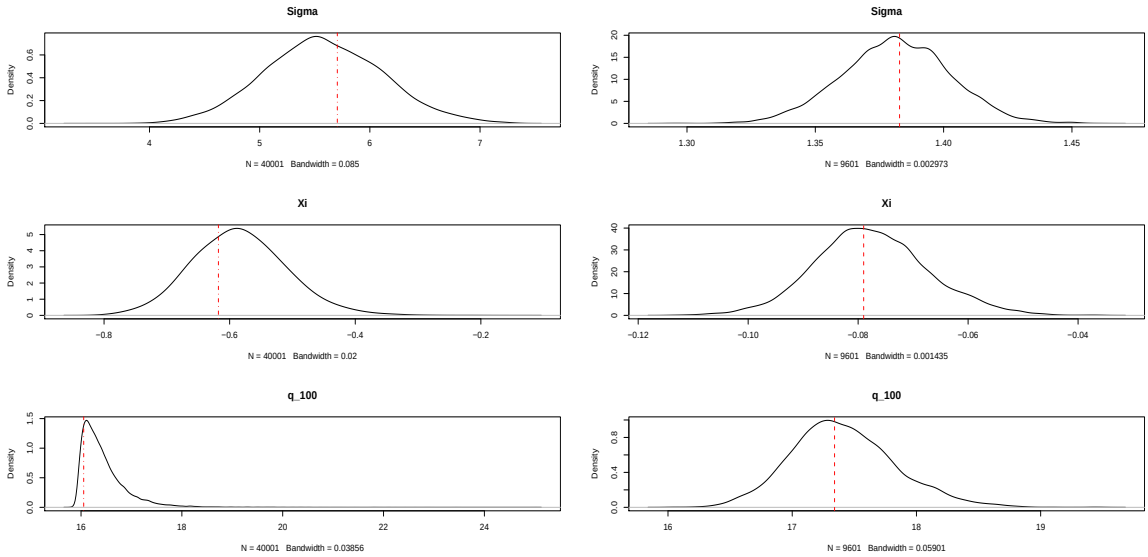

Figure 5.2: Sampling distrbutions for the posterior densities for cluster peaks and all excesses, respectively

Discussion

better bivariate method - point process model

References