Predicting the unpredictable: An Extreme Value analysis of rainfall data

Joseph Matthews

May 1, 2014

Abstract

Extreme weather events are becoming more and more frequent across the globe, causing devastation to buildings and human life. We attempt to model an example of extreme weather, namely rainfall from sites across Great Britain using the Generalised Extreme Value distribution, firstly from a frequentist perspective, then allowing for a trend in time with the location parameter. We then fit the GEV using a Bayesian MCMC methodology, again allowing for temporal trend in the location parameter, followed by assuming a constant trend and shape parameter across regions of sites. Finally we fit a random effects model to each parameter within sub-regions and compare in order to infer any patterns, before evaluating the strengths of our models using goodness of fit tests.

Contents

Chapter 1

Motivation

In this era of climate change and global warming, we are seeing more and more extreme weather events and environmental catastrophes, which are causing immense damage to infrastructure as well as ruining and claiming human lives in a variety of different countries throughout the globe. These can take many forms, for example hurricanes, the most pertinent in recent memory being Hurricane Katrina in 2005 which caused an estimated 81 billion dollars worth of damage and killed 1836 people [9], flash floods, such as those in Bangladesh in as recently as 2012 which killed 100 people and stranded a quarter of a million [1], and typhoons, such as Typhoon Haiyan in the Phillippines which killed more than 6000 people and forced over 3.8 million from their homes [8]. Of course these events are not restricted to the more tropical and exotic regions of the world, here in the UK we have also been affected by extreme weather, most recently during the winter storms in early 2014 which brought with them extreme winds and flooding, causing chaos throughout the country [5].

Naturally therefore it would make sense for us to attempt to predict these extreme events so as to better prepare ourselves when they strike and to allow us to construct appropriate safety mechanisms and protocol to protect both property/infrastructure as well as human life. Clearly this is not a simple task as these events can be dependent on a wide variety of factors, making exact predictions as to when they will occur extremely difficult, as well as the fact that we are seeing more and more that these events are in fact the most extreme in recorded history at that particular area, meaning essentially the task ahead is to predict that which has never happened before. It is at this point we make a compromise and state that whilst it would be useful to have a good prediction of exactly when a cataclysmic weather event will

Figure 1.1: Aftermath of Hurricane Katrina [7]

occur, in most cases it is correctly gauging the ferocity and scale of it which is of the most importance, since the primary use of this prediction will be, as discussed earlier, to construct appropriate defences against these disasters and so for this purpose it is important to know how strong the defences will need to be.

An example of this, which will provide the bulk of this report, is that of extreme rainfall which can cause severe flooding, and so it is important for the civil engineers responsible for the construction of sea walls etc. to prevent rivers from bursting their banks, to know how much rainfall to expect, and thus how high their wall will need to be. Herein lies a trade-off, since from a safety perspective one would argue it is best to simply build a sea wall as high as possible, although this will clearly lead to budgetary (amongst other) issues, meaning this is not practical and so a compromise must be reached. In order to reach an appropriate compromise we must have an effective method of forecasting the severity of the rainfall we are to expect, which is where the statisticians come in.

Chapter 2

Foundations of Extreme Value Theory

2.1 Background

Usual statistical methods in this area, such as those used to analyse the effects of global warming are based around looking at average levels over a given time period, or general patterns and trends in time. Clearly this standard statistical approach is of little use here, since it is not the average occurrences which cause the devastation discussed earlier, it is the most extreme events for which we must prepare. There are many approaches within Extreme Value Theory which attempt to solve this problem, a common feature among them being that we do not concern ourselves with the entire dataset, instead choosing to focus on the most "extreme" (usually but not necessarily the largest, depending on the event being measured) values and ignoring the rest. The method for defining which values we class as "extreme" varies between methodologies, that which we shall consider here involves simply dividing the time over which the observations were recorded into set intervals (e.g. days/weeks/years etc), and collecting the most extreme value from each interval.

2.2 Extremal Types Theorem

We now take a more abstract, general viewpoint of this act of observing extremes in order to understand the methods involved, before then showing how these can be applied to the rainfall dataset we shall be investigating. We suppose therefore we have a set of n i.i.d random variables $X_1, X_2, ..., X_n$ from a particular distribution which may or may not be known. We then denote the maximum of these variables as

$$M_n = max\{X_1, X_2, ..., X_n\} \tag{2.1}$$

where n is the number of observations. It is then this object M_n that we wish to learn about, in particular we would like to obtain a limiting distribution for it, as this would then allow us to make appropriate inferences on the maxima of our random variables. We can obtain an asymptotic result, known as the *Extremal Types Theorem* which states for given sequences a_n and b_n we have that

$$Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) \rightarrow G(x) \text{ as } n \rightarrow \infty \tag{2.2}$$

Where G(x) is one of the following three CDFs:

$$G(x) = \exp(\exp(-x)) - \infty < x < \infty \tag{2.3}$$

$$G(x) = \begin{cases} 0 & if \quad x \leq 0, \\ \exp(-x^{-\alpha}) & if \quad x > 0, \alpha > 0. \end{cases} \tag{2.4}$$

$$G(x) = \begin{cases} \exp(-(-x)^{\alpha}) & if \quad x < 0, \alpha > 0, \\ 1 & if \quad x \geq 0. \end{cases} \tag{2.5}$$

[3]

The above densities are all named after the people who worked most prominently with them, density 2.3 is known as the *Gumbel* density, density 2.4 is known as the *Frechet* and density 2.5 is known as the Weibull. We can see directly that the Gumbel density is unbounded, whereas the Frechet density is bounded below and the Weibull is bounded above. A plot of the 3 extremal densities is given in figure 2.1, where we can see that whilst they have similar shapes there is a clear difference in their curvature, this is accentuated further with larger values of α in the Frechet and Weibull densities (we show a relatively modest $\alpha = 1$ in 2.1)

We elect to give an outline proof of this result here but first must define the property of max-stability as follows:

A distribution G is max-stable if for all $n \geq 2$, there exists constants $\alpha_n > 0$ and β_n such that

$$G^n(\alpha_n z + \beta_n) = G(z) \tag{2.6}$$

Recalling our definitions from earlier in the chapter where G(z) is the CDF of $M_n = max\{X_1, X_2, ..., X_n\}$, we see that equation 2.6 simply means that if

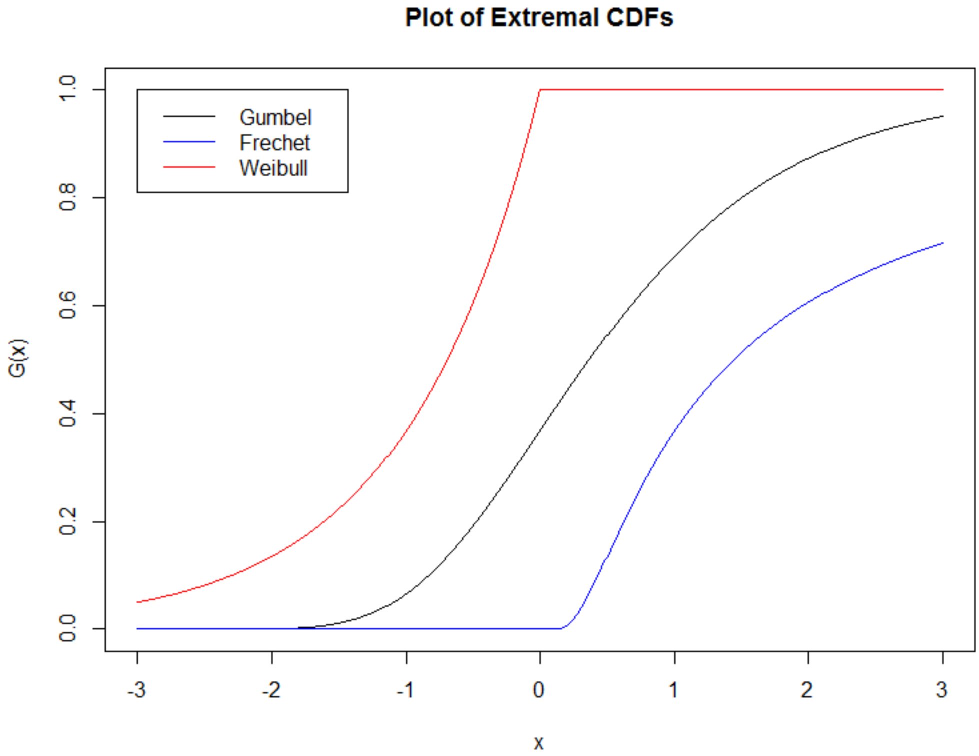

Figure 2.1: Plot of extremal distributions

our density G(z) is max-stable, then it will remain the same type of extremal density (that is Gumbel, Frechet or Weibull) irrespective of how many maxima we are using in our sample, and thus will be identical up to changes in location and scale parameters.

From this definition we propose that a density is max-stable if and only if it is a member of the extremal family of distributions outlined earlier. We can see with simple algebra that the distributions defined above (2.3,2.4,2.5) are max-stable, and the converse can be proved, although this requires advanced functional analysis and so we do not state the proof here. Hence in order to prove the Extremal Types Theorem we simply have to show that $M_n k$, the maximum random variable in a sequence of $n \times k$ random variables (either the maximum of a single sequence of length $n \times k$ or the maximum of k maxima, each of which is the maximum of n observations) is max stable (and hence a member of the GEV family).

We suppose the limit distribution of $\frac{M_n-\beta_n}{\alpha_n}$ is G, i.e. for large n $Pr(\frac{M_n-\beta_n}{\alpha_n} \leq z) \approx G(z)$. Hence for any k (with large n) we have

$$Pr\Big(\frac{M_{nk}-\beta_{nk}}{\alpha_n} \leq z\Big) \approx G(z) \tag{2.7}$$

However we defined M_{nk} to be the maximum of k variables which have the same distribution is M_n, and so:

$$Pr\left(\frac{M_n-\beta_n}{\alpha_n} \leq z\right) = Pr\left(\frac{M_n-\beta_n}{\alpha_n} \leq z\right)^k \tag{2.8}$$

Hence by equation 2.7 we have:

$$Pr(M_{nk} \leq z) \approx G\left(\frac{z-\beta_{nk}}{\alpha_{nk}}\right) \tag{2.9}$$

and by equation 2.8 we have:

$$Pr(M_{nk} \leq z) \approx G^k\left(\frac{z-\beta_{nk}}{\alpha_{nk}}\right) \tag{2.10}$$

Hence we have that G and G^k are the same apart from the sequence of a and b values (the scale and location coefficients) and hence we have that G is max-stable and therefore is a member of the GEV family ([2, p. 49-51]).

This is a powerful result, since it shows that the maxima of random variables from any distribution will have CDF given by one of the above distributions. Indeed it is worth noting that in each of equations equations 2.3, 2.4, 2.5 that the initial distribution of the random variables does not appear at all.

We consider the Extremal Types Theorem to be almost analogous to the Central Limit theorem, in that the sequences of values a_n and $b_n$behave like the standard deviation and mean do in the Central Limit Theorem, that is they are essentially normalising values. We do not concern ourselves with their values however since they are just constants and so can be absorbed into the model, and so we shall not make any further mention of them here.

2.3 GEV distribution

Clearly we have already achieved a powerful result with the Extremal Types Theorem in section 2.2 providing us with just 3 possibilities for the distribution of our maxima, however we naturally seek to generalise yet further and obtain just a single distribution to which all maxima can be attributed, this is known as the Generalised Extreme Value distribution (henceforth referred to simply as the GEV) which is given as follows:

$$G(x;\mu,\sigma,\xi) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]_{+}^{-\frac{1}{\xi}}\right\} \tag{2.11}$$

However clearly this is undefined in the limit when ξ tends to zero, in this case we use the Gumbel approximation, given in equation 2.12

$$G(x;\mu,\sigma) = \exp\left\{-\left[\frac{x-\mu}{\sigma}\right]\right\} \tag{2.12}$$

[4] & [6]

From 2.11 we see that we now have 3 parameters governing this distribution, firstly $\mu \in (-\infty,\infty)$ which is known as the location parameter, and in keeping with the Normal distribution comparison made earlier, this behaves similarly to the mean of the Normal, in as much as it determines the position of the density on the number line, and the general magnitude of the values contained (although it should be made clear that μ is not the mean of the GEV). Next we have $\sigma \in [0,\infty)$ which is the scale parameter, and again behaves similarly to the variance/standard deviation of the Normal in as much as it determines the spread if the density (although again it is not the variance nor the standard deviation of the GEV). Finally we have $\xi \in (-\infty,\infty)$ which is the shape parameter, which as its name would suggest, determines the shape of the GEV density. In particular it determines which of the previous 3 densities (equations 2.3, 2.4, 2.5) the GEV will take the shape of, according to the following set of criteria. If $\xi = 0$, then the density is a Gumbel, if $\xi < 0$ then the density is a Weibull, and if $\xi > 0$ the density is a Frechet. With the exception of the former this can be seen directly by comparing 2.11 with the corresponding density from section 2.2 (equation 2.4 or equation 2.5) with positive/negative values of ξ.

We can verify this property of ξ graphically by comparing figure 2.2 with figure 2.1 from section 2.2. We see that the curve with $\xi = -1$ displays the same quadratic like curvature as the Weibull density, similarly the $\xi = 1$ curve displays the same levelling off type curvature as the Frechet density. We recall that when $\xi = 0$ the GEV is undefined and hence is replaced by the Gumbel density, and so there is nothing to verify in this case. Furthermore, it is also worth noting that whilst theoretically ξ can take any real value, it is rare to see values outside of the range $(-1,1)$, and to do so usually indicates that the model has not fitted the data well for some reason, and so we included the most extreme cases we are likely to encounter in figure 2.2.

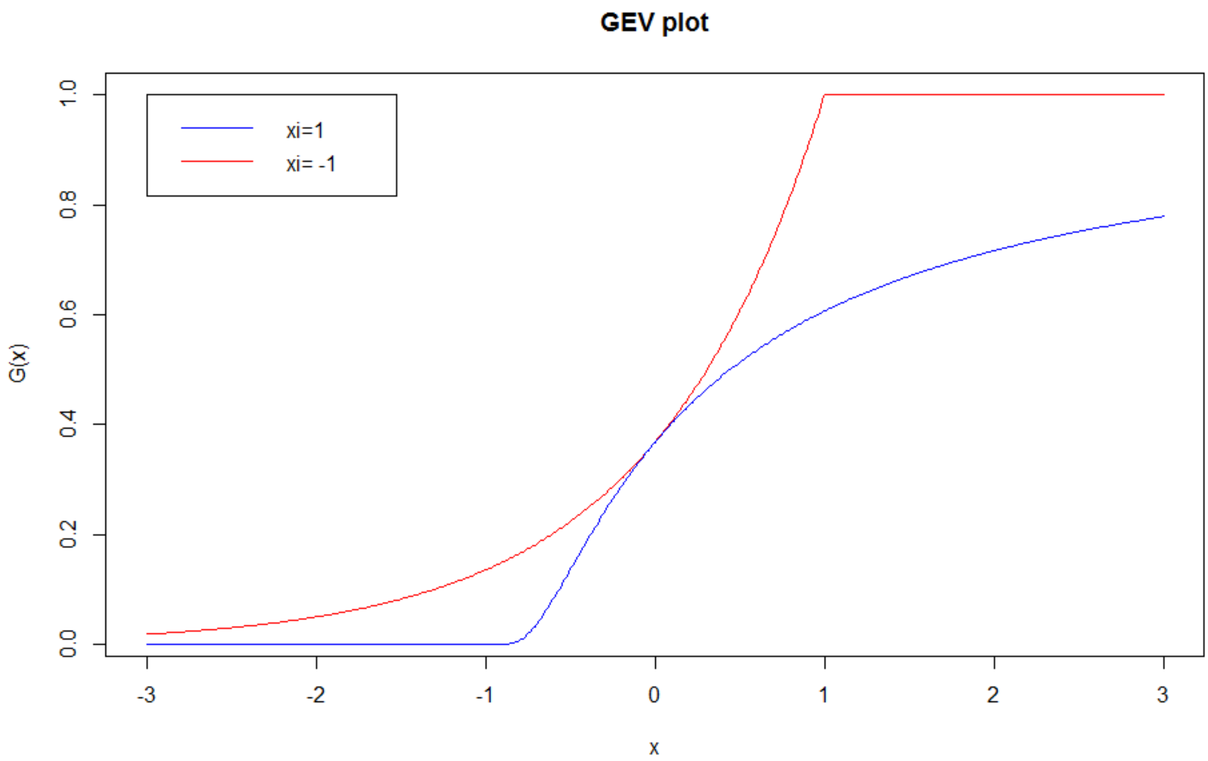

Figure 2.2: Plot of GEV CDF

2.4 Applying the GEV

Now that we have obtained a distribution which we can use to model extreme values, the natural question is then to ask how do we use this model to provide inference which can help in the real world to solve the problems discussed in chapter 1? Since we have already obtained the CDF of the GEV, the first step may be to calculate exceedance probabilities, that is to determine the probability that the next maximum will exceed a given value. Whilst this is a perfectly valid step to take and can be used to obtain some useful results, it is not the most prevalent usage of GEV models, instead it is far more common to calculate *return levels*.

Return levels are an inverted form of the exceedance level calculation, where then we were interested in obtaining a probability of exceeding a fixed value, whereas now we fix the probability and the return level is the value which shall be exceeded with that probability. However, within the context of extremes we are rarely, if ever, simply concerned with what shall be happening in the immediate future, but instead need to think more long term. With this in mind we choose to interpret these probabilities, not as "the probability the next maxima shall be higher than a given value", but instead we interpret it as the relative frequency with which that value shall be exceeded. For example, if we were to obtain an exceedance probability of 0.01, we could interpret this as a fraction 1/100, and conclude that we would expect to see this value exceeded once in the next 100 maxima. Similarly we therefore interpret return levels as finding the value we would expect to be exceeded once in a given length of time. It is important to bear in mind when taking this interpretation of return levels however that they are based on the current values of the parameters of the GEV distribution remaining constant, and so, later on when we come to allow the parameters of our GEV density to vary in time, our return levels will be based on the parameter values for the year from which we are calcluating the return level. We obtain the formula for return levels by inverting the GEV density (equation 2.11) to make x the subject, and in doing so obtain

$$z_r = \mu + \frac{\sigma}{\xi}\left[\left(-\log\left(1-\frac{1}{r}\right)\right)^{-\xi} - 1\right] \tag{2.13}$$

We observe that this density, like the GEV density from which it is derived, is not defined for a Gumbel type distribution (where $\xi = 0$) and so for this case we invert the Gumbel density (equation 2.3) to obtain:

$$z_r = \mu + \sigma\left[\left(-\log\left(1-\frac{1}{r}\right)\right) - 1\right] \tag{2.14}$$

Here we have the formulae for what is referred to as the r-year return level (that is the value which is exceeded with probability 1/r) for a given set of GEV parameters. It can be seen directly that the value of a fixed r-year return level increases with μ, which is perhaps obvious when we consider the role of μ within the distribution, whilst it decreases with increasing σ and ξ.

2.5 The Dataset

We now consider the dataset which we shall be studying throughout this report. We have hourly rainfall measurements, given in tenths of millimetres, taken over 62 years at 1281 separate sites across Great Britain (although for each site some of the entries are missing) which are divided in regions, and then further into sub-regions geographically.

We have several problems with using the hourly data as it stands though, the perhaps most important, is a mathematical issue arising from section 2.2, when we derived the Extremal Types Theorem, one of the requirements was that our data points we independent identically distributed random variables. Clearly this would provide problems since we cannot realistically assume that the amount of rain falling in each hour of the day will be from the same distribution across all days (which is made obvious when we consider seasonal differences etc) and also the amount of rainfall in an hour is highly unlikely to be independent of the amount of rain in the surrounding hours (an hour where we observed 10mm of rainfall is highly unlikely to be surrounded by hours where no rain fell at all), and so we have a problem.

Fortunately there is a relatively simple way to solve this issue. We choose to transform our dataset by instead of working with the hourly observations, we sum each day to obtain a set of daily totals for each site, and we now need to select a timeframe from which the maximum daily totals will be independent and identically distributed. Clearly selecting weeks or months would be unlikely to be valid since there would be clear seasonal differences still present, however this is not true if we were to take the largest value per year (referred to as the annual maximum). We can be satisfied that observations from different years are sufficiently far apart that they can be considered independent, and since we have removed any seasonal effects etc we can be satisfied that they would be from approximately the same distribution, and so our assumptions for the Extremal Types Theorem are satisfied, when we consider the maximum daily totals per year.

Whilst our dataset is incredibly useful with regards to inference, given it's extraordinary depth, we do have the issue of each site having large numbers of entries missing for whatever reason (monitoring stations having not been built, closed down, temporarily out of commission etc). This not only provides the obvious disadvantage of our dataset not being as large as it could be, it can also endanger some of the key assumptions on which our modelling approach is based. Given that we will be simply extracting the most extreme day of each year, we do not know how many days of that year were missing, and so clearly we will run into problems if we treat 2 maxima the same if one was from a year with 365 of recorded days and the other with 10. For this reason we impose a minimum bound on the amount of data that must be recorded in order to make a year "valid" and thus OK to include in our model. Since the data is hourly it is easier to simply impose the limit on the number of recorded hours, and so we choose our lower bound to be 7300 hours (out of a possible 8760, i.e. 5/6) of data must be recorded for the year to be accepted. Whilst this is an arbitrary choice (selecting 7000 or 8000 for instance would not cause dramatically different results), it does provide a robust set of results which are unaffected by the fact we have missing data, although regrettably this will lead to a loss of data in our model.

Chapter 3

A Frequentist Approach

3.1 Maximum Likelihood Estimation

We have now established all of the basic theory required in order to make some good inferences regarding extreme value estimation, and must now turn our attention to actually applying this theory to our dataset. From a frequentist perspective our first instinct when fitting a model like this to a dataset is to use maximum likelihood theory to obtain estimates for each of the parameters. As normal we obtain an expression for the likelihood of the GEV from its density, obtained by differentiating 2.11 to obtain:

$$f(x;\mu,\sigma,\xi) = \frac{1}{\sigma} exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right] exp\left\{-exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\} \tag{3.1}$$

Which hence gives:

$$\begin{aligned} L(x;\mu,\sigma,\xi) &= \prod_{i=1}^{\infty} f(x_i;\mu,\sigma,\xi) && (3.2) \\ &= \prod_{i=1}^{\infty} \frac{1}{\sigma} exp\left[-\left(\frac{x_i-\mu}{\sigma}\right)\right] exp\left\{-exp\left[-\left(\frac{x_i-\mu}{\sigma}\right)\right]\right\} && (3.3) \end{aligned}$$

From this we can then obtain an expression for the log-likelihood. As with the return level formulae we have algebraic problems in the case $\xi = 0$ (i.e. with a Gumbel density) and so this must be defined separately. When $\xi \neq 0$, we have:

$$l(x;\mu,\sigma,\xi) = -n\log(\sigma) - (1+\frac{1}{\xi})\sum_{i=1}^{m}\log\left[1+\xi(\frac{z_i-\mu}{\sigma}\right] - \sum_{i=1}^{m}\left[1+\xi(\frac{z_i-\mu}{\sigma}\right]^{-\frac{1}{\xi}} \tag{3.4}$$

And for $\xi = 0$ we have:

$$l(x;\mu,\sigma) = -n\log(\sigma) - \sum_{i=1}^{m}\frac{z_i-\mu}{\sigma} - \sum_{i=1}^{m} exp\left\{-\frac{z_i-\mu}{\sigma}\right\} \tag{3.5}$$

We then wish to obtain parameter values which maximise these likelihood functions, unfortunately we do not have closed form for these maximum likelihood estimators, and so we compute our maximum likelihood estimates via numerical methods. Fortunately this can be done easily within the "ExtRemes" and "ismev" packages in R. We then use these maximum likelihood estimates to make inferences about the extremes at the site in question, the most obvious examples being the estimate of the shape parameter gives us whether the distribution is of type Gumbel, Frechet or Weibull, as well as allowing us to determine return levels at the various sites, for any given year- usually 50 and 200 are popular choices for this inference, and indeed using a 50 year return level is a legal requirement for civil engineers when constructing weather defences.

Now that we have established the methodology, we can begin applying the model and obtaining some output. Using the previously mentioned "ExtRemes" package in R we apply this model to the "Heaton Park" site (code given in Appendix .1) and obtain the following output:

$$\hat{\mu} = 298.84,\ \hat{\sigma} = 61.94,\ \hat{\xi} = -0.50,$$
$$s.e.(\hat{\mu}) = 17.94,\ s.e.(\hat{\sigma}) = 14.15,\ s.e.(\hat{\xi}) = 0.23$$

By direct substitution into 2.13 we can therefore obtain values for return levels, i.e. $\widehat{z_{50}} = 405.11$ and $\widehat{z_{200}} = 413.94$.

The main inferences we can make here are that surprisingly the rainfall from this site took a Weibull density (since $\xi < 0$), and upon comparison with its standard error we see that it is greater than 2 standard errors less than 0, meaning we can be confident that this is not due to random chance (although how representative this particular site is remains to be seen). Our return levels show that the largest amount of daily rainfall we would expect to see at Heaton Park in the next 50 years is 40.51mm, and similarly the largest in the next 200 years is 41.39mm. We also obtained a negative log-likelihood value of 81.4027, which can be used for simple goodness of fit tests, which shall be referred to shortly.

3.2 Additional Parameters

One fairly large limitation of this approach is that it assumes the process is stationary, that being that the levels of extreme rainfall are not changing with time, shown by the fact that the model parameters are constant. We can investigate this assumption by allowing the parameters to vary in time; whilst theoretically we could allow all of our parameters to vary, here we choose just the location parameter because as mentioned in section 2.3, the location parameter governs the general position of the distribution, and thus the overall magnitude of the extremes, and hence allowing this parameter to vary would reflect the changing magnitude of the extreme weather events. Hence we are effectively setting $\mu = \mu_\alpha + \mu_\beta * time$, where μ_α is the intercept term and μ_β is the effect of time, known as the "trend" or "slope" term. Refitting this model to the site shown previously (using code given in Appendix .2) gives:

$$\hat{\mu_\alpha} = 293.52,\ \hat{\mu_\beta} = 24.11,\ \hat{\sigma} = 51.15,\ \hat{\xi} = -0.26$$

$$s.e.(\hat{\mu_\alpha}) = 16.97,\ s.e.(\hat{\mu_\beta}) = 21.34,\ s.e.(\hat{\sigma}) = 14.14,\ s.e.(\hat{\xi}) = 0.38$$

$$\widehat{z_{50}} = 455.60,\ \widehat{z_{200}} = 479.57$$

From this we can see there is a positive trend with time, with our fitted estimates indicating the location parameter increases by around 24 year-on-year, however upon comparison with the standard error for the slope parameter, we see that it is only just greater than 1 standard error greater than 0, which implies that this result alone is not sufficient to conclude that this trend exists, and would need to be combined with other data to do so (as we shall see shortly). We have also obtained a Negative log-likelihood value of 80.6490 for this model, again which we shall use shortly. In comparison with the model fitted in section 3.1 we see that the value of the intercept term μ_α is lower than the original μ estimate, which is to be expected since μ has effectively been divided between two positive parameters. We also note that the values of σ and ξ have changed slightly with the addition of the new term, and interestingly the standard error for ξ is now greater than its value (in modulus) meaning we can no longer be confident that its true value is indeed below 0 (and thus that the density is a Weibull). We also observe that our 50 and 200 year return levels are now both higher than

their counterparts when trend wasn't included, which implies that allowing for this positive trend throughout the duration of the data has lead to expect more severe weather events in the future, as opposed to when we assumed the parameters were remaining constant throughout the data.
Whilst we have made a preliminary analysis on the effectiveness of the slope term using its standard error, there is a more robust method. Since we have a pair of nested models (since they are identical with the exception of an additional parameter) we can therefore carry out a likelihood ratio test to determine whether the accuracy of our model has been significantly improved with the addition of μ_β, which then provides information as to whether the temporal trend really exists in any significant way. As seen in the above results we have obtained the negative log-likelihoods for each of our models, we can use these to calculate our test statistic:

$$D = 2 * (81.4027 - 80.6490) = 1.5074 \tag{3.6}$$

Under the null hypothesis (that there is no difference between the models, i.e. there is no significant trend) this value has a chi-squared distribution with one degree of freedom (since the difference in numbers of parameters between the models is 1), and we can use R to the p-value as $p = 0.220$. As ever, since $p > 0.05$ it is not significant, and so we conclude we have no evidence to reject the null hypothesis, and hence have no evidence to support there being a temporal trend in the location parameter at this site, which is in support of the primitive inference we made earlier with the standard error.

Chapter 4

A Bayesian Approach

4.1 Markov Chain Monte Carlo

We have shown in chapter 3 how to fit a standard GEV model using frequentist methods, namely maximum likelihood estimation. Whilst this is a perfectly valid approach, in modern statistics it is usually preferable to attempt a Bayesian analysis instead, as this holds many advantages when it comes to interpreting the results we obtain.

With a distribution as complicated as the GEV, it is regrettably not possible to use a simple closed form method such as a conjugate analysis to carry out our analysis, however it is possible to carry out an MCMC method to derive our posterior densities, which in algorithm which generates values which generates values for our posterior, which we then decide whether or not to accept it based on how well it fits to the data. A key question we must consider when conducting a Bayesian analysis is what prior distributions do we use to represent our beliefs regarding the parameters? Since we have no expert knowledge to influence our beliefs prior to fitting the model, we have to be wary of selecting a prior distribution which will heavily influence our posterior distribution, since we have no prior knowledge, we essentially wish for the data itself to dictate our conclusions. In order to do this we select "vague priors" which are essentially prior distributions with very high variances, to reflect the fact that we have no prior inclinations regarding the parameter values, and so theoretically they could be anything. The vague priors are Normal $N(0, 1000^2)$ distributions for each parameter. For our update mechanism we use a metropolis random walk with Normal innovations, with mean 0 and variance tailored to the parameter in question in order to ensure the chain mixes well (we shall describe what this means later). Since the scale parameter is constrained to be strictly positive, in order to facilitate these normal innovations (which by definition can be negative), we choose to work with the log of the scale parameter, which we shall denote η, which allows us to take negative values, whilst ensuring strictly positive values for the scale.

We must also select the starting states for our chain, again these are tailored to the individual parameters. For the location parameter we choose to initialise at the mean of the data we are fitting our model to (namely the series of annual maxima for the site in question) and the scale parameter at the log of the standard deviation of the data, this is since, as discussed previously, the location and scale parameters play similar roles to the mean and standard deviation of the distribution, and so therefore it would seem logical that our final parameter estimates would be similar to the mean and standard deviation provided by the data, particularly since we are using vague priors which should not influence the posterior densities.

Finally we choose to initialise the shape parameter at 0.01, ideally we would select it to start at 0 since it can take both positive and negative values, this would require us to switch to the Gumbel form of the likelihood, and so it is simpler computationally if we are to stick to the GEV likelihood where possible, and so we select a very small non-zero value instead. Whilst we have tried to be logical with our choice of starting value, by allowing our chain to run for a long time the starting values should become irrelevant as the chain settles down to its stationary distribution (this can be checked by analysing trace plots of the parameter values as the chain progresses, which we shall look at later). We now apply this Bayesian model to the Heaton Park site we looked at in chapter 3. As discussed above it is important to tune the Normal innovations so that the chain mixes well, and hence gives sensible output, this can be checked by studying the "acceptance rate" (the proportion of the time that the value of a parameter moves, i.e. a proposed value is accepted), with rates of around 0.3 indicating that the chain is mixing well. Ultimately innovation values of 30, 0.8, and 0.45 for μ, σ and ξ respectively, gave rates of 0.435, 0.266 and 0.316, which we deem acceptable. The mixing of a chain can also be investigated by viewing plots of the parameter values as the chain progresses, shown in 4.1.
In figure 4.1 we can see that the value of each parameter changes fairly frequently and for the most part provides the dense black area which implies it is mixing well (although admittedly not perfectly, as there are times when the value remains stuck in the same place for a noticeable period, however this is infrequent and so we can look past it). In addition the parameter

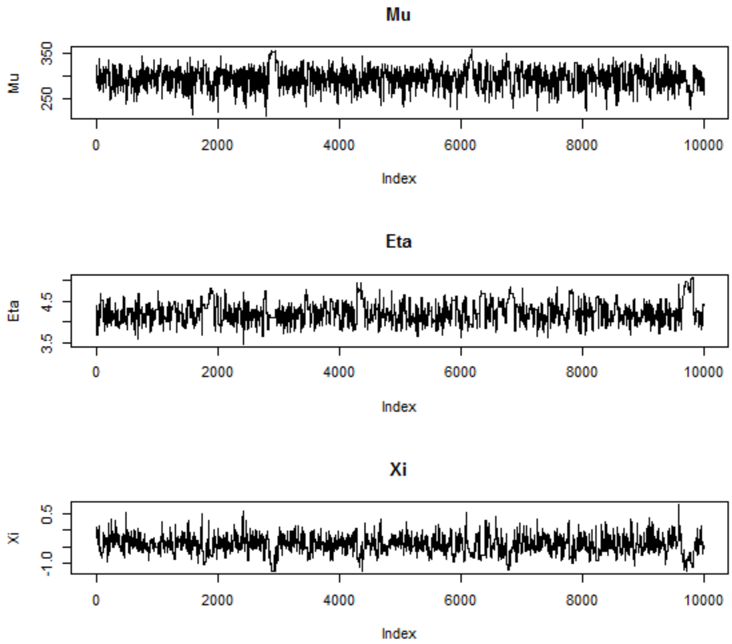

Figure 4.1: Traceplots of parameters

values remain within the same area, which implies the chain is mixing well, and that it has converged to the stationary distribution (i.e. the range of values that are being accepted by our chain are those which come from the posterior distribution). We can then view this obtained posterior distribution by constructing histograms of the parameter values throughout the chain, shown below in figure 4.5.
In figure 4.2 we can see the distribution of each parameter is fairly symmetric around a single modal peak, in this case we see μ peaks at around 300, σ peaks at around 60 (after exponentiating the chain values since the parameter in the chain was $\eta = \log(\sigma)$) and ξ peaks at around -0.5. This is typical of what we would expect in an ideal situation, however when models become more complex the histogram may not be as easy to interpret as this. If we were to compare this output with the frequentist output obtained in section 3.1 we can see that our modal peak roughly coincides with the maximum likelihood estimates for this model which we calculated to be $\hat{\mu} = 298.84,\ \hat{\sigma} = 61.94,\ \hat{\xi} = -0.50$. In addition to this the histogram shows that the majority of the values we accept for μ to be in the region of (250,350) which by and large coincides with the maximum likelihood estimate for the standard error obtained in section 3.1 of 17.94,and so if we apply the rough estimate of expecting most of the data to lie within 2 standard errors of the

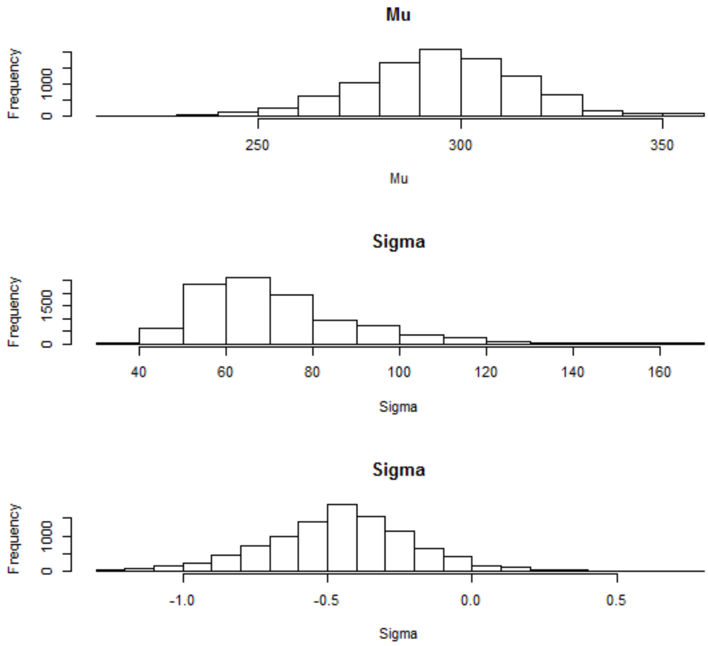

Figure 4.2: Histograms of posterior densities

mean this would give an interval of roughly (264,336). Similarly the majority of the σ values are in the range (40,100), which ties in with it's standard error being around 15, and the ξ values tend to lie in the approximate area (-0.8,-0.3), which ties in with the standard error being around 0.25. Hence the majority of our Bayesian inference gives results roughly similar to their frequentist counterparts from section 3.1 which is not surprising given the vague prior distributions we are using in our model, designed to allow the data to dictate the results.

A further useful property of the Bayesian MCMC method for fitting the GEV is the use of return levels, since with the frequentist approach we only obtain a single return level based on the maximum likelihood estimates for the GEV parameters, here we can obtain a return level for each iteration of our chain, and hence develop a distribution for our return levels. The 50 and 200 year return levels for this model are given in figure 4.3.

In figure 4.3 we observe that like the standard histograms shown in figure 4.2, our return level densities have a clear modal peak, and the remaining density appears to be roughly symmetrically distributed around it. We observe that for both the 50 and 200 year return levels the modal peak is around 400 for both, which tie in with the frequentist estimates in section 3.1

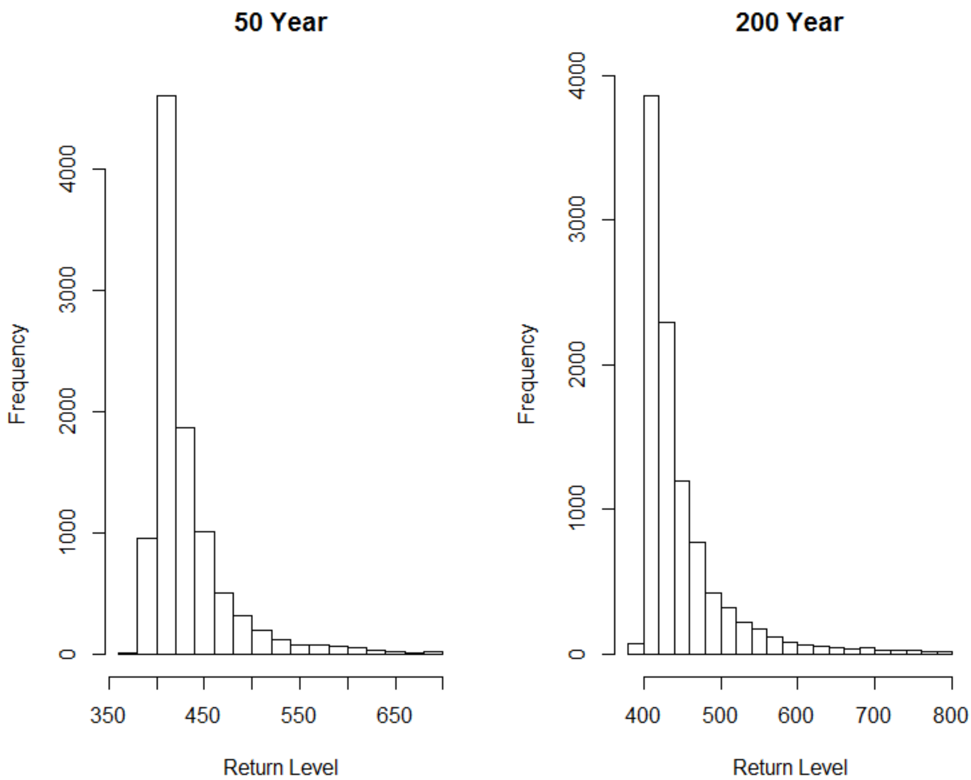

Figure 4.3: Histograms for return levels

of 405.11 and 413.95 respectively with the 200 year density shifted slightly to the right of the 50 year density, as we would expect.

4.2 Fitting Trends

As before with the frequentist approach in section 3.2 we would quite like to investigate whether our parameters, particularly the location parameter, vary in time. This can be done quite easily by replacing all instances of μ in our model (including likelihood etc) with $\mu_\alpha + \mu_\beta * time$ and then treating the new slope parameter exactly the same as the others in our model by including it within the random walk metropolis scheme used for the other parameters. Again we do not have any information on which to base our prior distribution for this new parameter, and so we adopt the same vague prior mechanism as with the others, selecting a Normal $N(0, 1000^2)$ prior. We tune the variance of the Normal innovations in this scheme in the same way as for the others, and elect to use a starting value of zero, which is a purely arbitrary choice as we have no idea as to what the final value will be, and again can rely on the length of the MCMC chain to remove the importance of this choice anyway. For numerical purposes, as in section 3.2 we standardise the year values so as to ensure the the slope term μ_β mixes well.

As before we tune the innovations to ensure a good mixing of the chain, ultimately selecting innovation values of 35, 35, 0.75 and 0.5 for μ_α, μ_β, σ and ξ respectively, as they yielded respective acceptance rates of 0.4017, 0.4526, 0.2509 and 0.3809. Again we can attempt to verify a good mixing of the chain by studying the traceplots in figure 4.4.

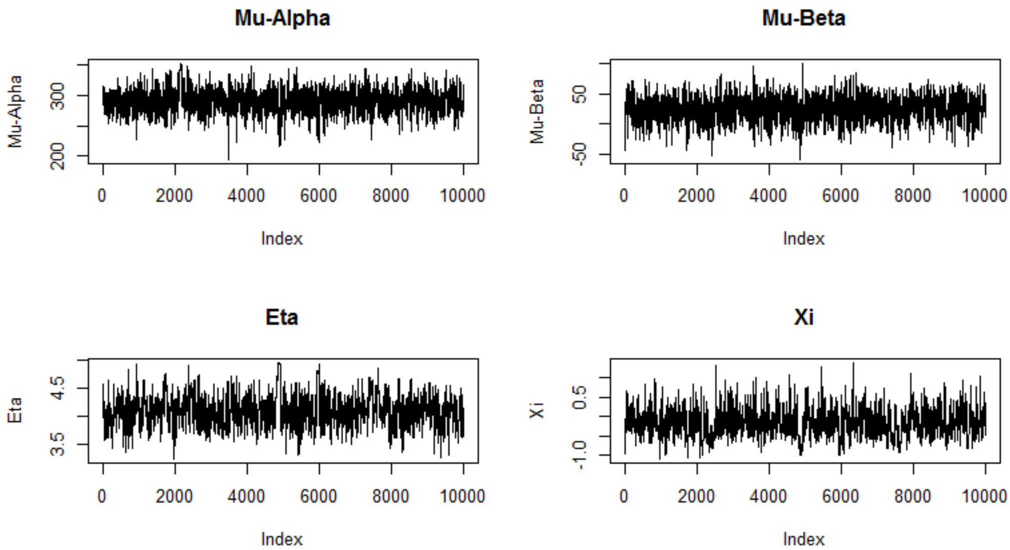

Figure 4.4: Traceplots of parameters

In figure 4.4 we see that once again each trace plot displays a thick black line, indicating that the chain is mixing well, and furthermore since the values remain within a single band, we can be satisfied that the chain has converged to its stationary distribution, and thus the values that we are sampling from are indeed the posterior values for each parameter. Again, we investigate this posterior using histograms of simulated values in 4.5.
In figure 4.5 we observe a similar result as in section 4.1 in as much as the Bayesian results here very much mirror their frequentist counterparts in section 3.2. We have singular modal peaks for each parameter as in 4.1 with μ_α peaking at around 300, compared with it's MLE estimate of 293.52, μ_β peaks at around 25, compared with 24.11, σ peaking at 50, compared with

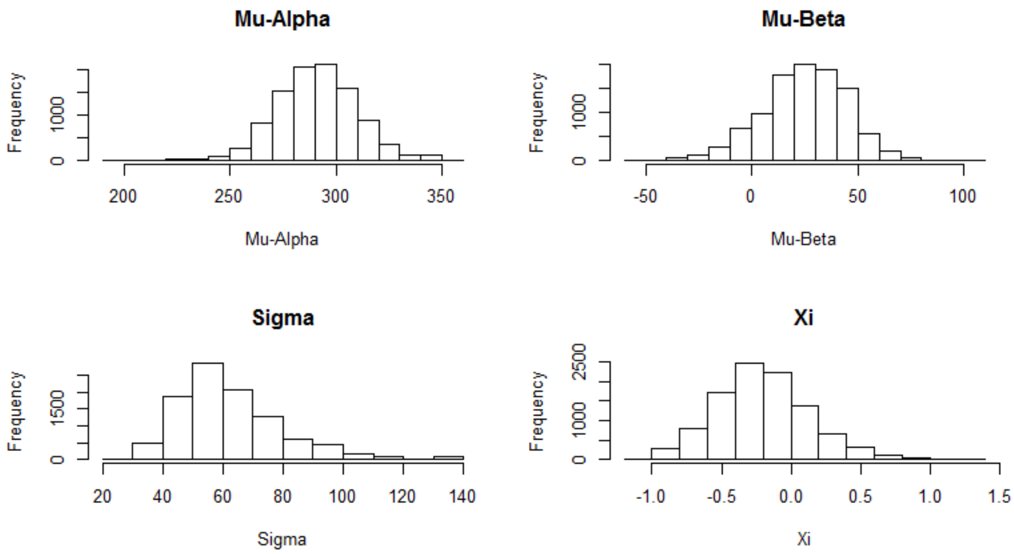

Figure 4.5: Histograms of parameters

51.15, and finally ξ peaking at -0.3 compared with -0.26. From the above list we again support our conclusions from section 4.1, that the Bayesian analyses generate very similar results to their frequentist counterparts. As in section 4.1 we can investigate the return level distribution, shown in figure 4.6.

In figure 4.6 we again observe a clear modal peak just above 400, which ties in well with the frequentist estimates in section 3.2 of 455.60 and 479.57, further supporting our beliefs that a vague prior approach to a Bayesian analysis yields similar results to a frequentist approach. It further supports the conclusions we forged in section 3.2 that allowing for trend causing an increase in return level values.

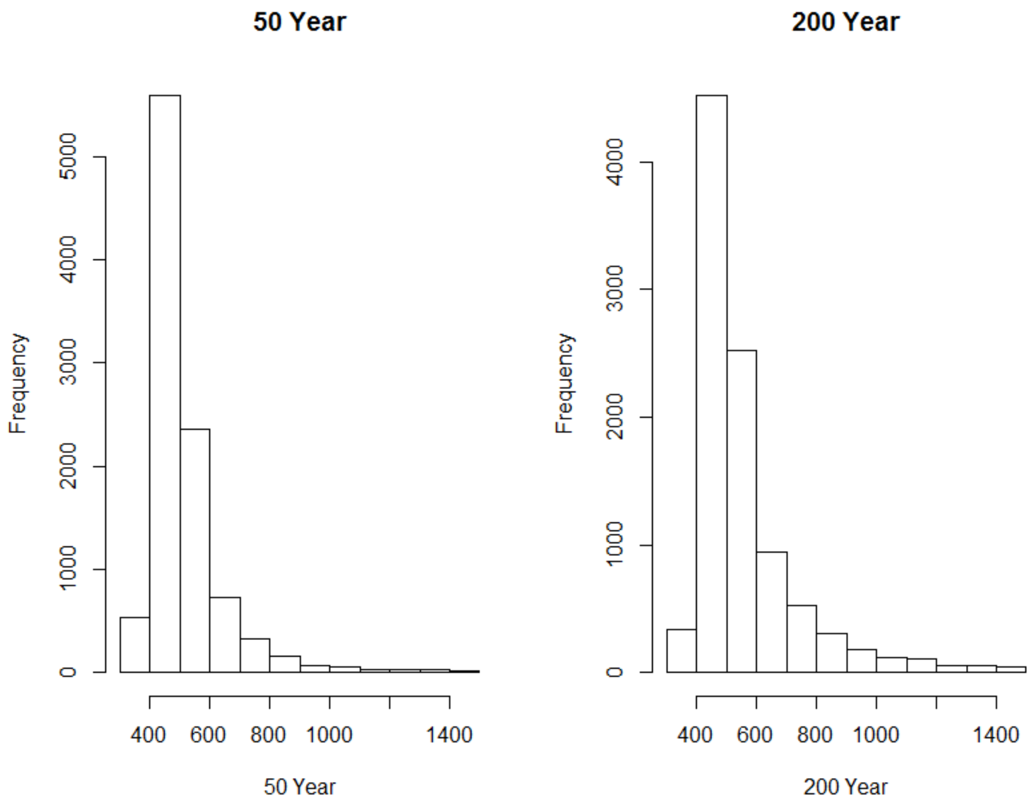

Figure 4.6: Histograms of return levels

4.3 Combining Parameters

Thus far we have developed a model which fits a seperate set of parameters to each site independently of data observed at surrounding sites within the same sub-region which in turn assumes that extremes at a particular site are completely independent of those in the surrounding area, which unlikely to be a realistic viewpoint. One particular area in which we can assume some solidarity among neighbouring sites is for the shape parameter ξ, clearly we would not expect the entire shape of the extremal distribution to change within the same small area, and so we take the step to assume a constant shape for all sites within a sub-region. Furthermore it would appear to be sensible to also assume a constant slope parameter μ_β within the region, as again it does not seem logical for the effect of time to have a dramatically different effect on the rainfall of regions which are quite close together. There are several advantages to this approach, chief among them being that it appears to be a realistic assumption, and in making it we greatly increase the amount of sata we have to support our inference regarding the values of ξ and μ_β (namely that we will be using the combined data for all of the sites in a sub-region for learn about one parameter, as opposed to just a single site as previously). Hence in order to fit this model we can no longer fit a model to a single site by itself as previously, and instead must fit to an entire sub-region at a time. For illustrative purposes we choose not to demonstrate this on a sub-region included in our main dataset, but instead on a dataset of pre-obtained maxima from Central and Eastern England, which is a set of 26 annual maxima (with no missing values) obtained from 21 sites, and for computational time constraints we reduce the number of iterations of our chain from 10000 to 2000, however we can be satisfied this is still sufficient for the chain to converge well, as is shown in the selection of traceplots (there are 63 in total) are shown below. The code for this model is given in Appendix .5.

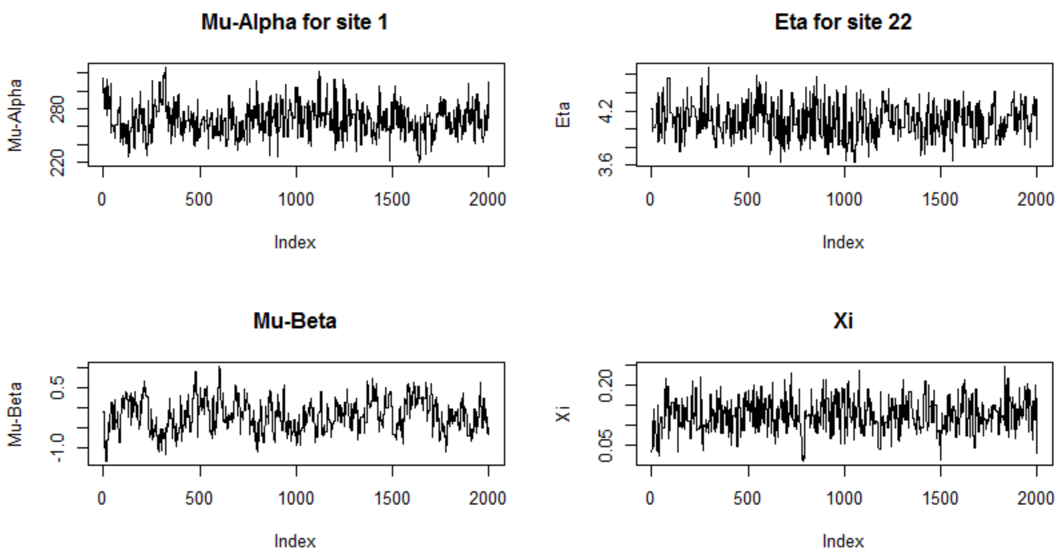

Figure 4.7: A selection of traceplots of parameters

From the traceplots in figure 4.7, we can see that whilst we do not have a dense black region as we would have liked in each plot, the parameter value does move frequently and occupies a good region of the space, and hence the lack of density in the line is most likely due to the reduced number of iterations in this run due to computational constraints. We do observe in addition that the most satisfactory tace plots are those of the combined parameters, ξ and μ_β, which is clearly due to the much larger dataset we had on which to base our model for these parameters. Since each of the trace plots occupy a band throughout the run of the chain we can be satisfied that it has converged as required and move on to study the histograms in order to make inferences regarding the posterior.

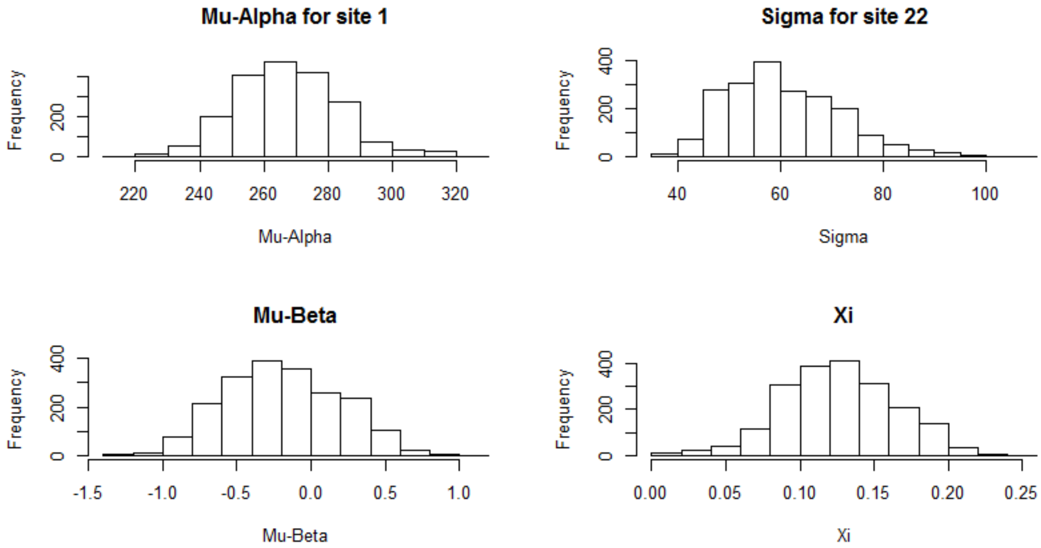

Figure 4.8: A selection of histograms of parameters

In figure 4.8, we observe that the combined parameters, in this case μ_β and ξ have a much smaller spread of values than for when they were simply for an individual site, again due to the fact we now have a much larger dataset on which to base our model for them. We also observe that from the histogram for the combined slope parameter for the entire site, the density is spread around zero, with both positive and negative values having significant density. Based on this result alone we would have to conclude that there is no significant evidence for a temporal trend being present at sites in this particular region.

The intercept and scale parameters are more likely to vary between these sites and so we choose not to combine parameters for these (not least since this would mean fitting a single GEV model to an entire sub-region, thereby ignoring any variability that is present).

Chapter 5

Random Effects

Throughout Chapter 4 we experimented with the idea of having distinct parameters for each site within a sub-region, and then in section 4.3 considered the opposite idea of combining certain parameters which we believed would be unlikely to change in any significant way within the same small geographical area (i.e. within the same sub-region). Whilst there is an argument supporting both methodologies, both also have their drawbacks. The former ignores the data we have on the surrounding sites, and the later does not allow for any variation at all between sites within the same sub-region-hence it would appear that a compromise between the two would be desirable. The compromise we use here would be to accept there are differing values of parameters within the same sub-region, however rather than treat them as completely unrelated- known as fixed effects- we treat them as realisations from an overall distribution from which the values of a particular parameter for each site in a sub-region are obtained- which are known as random effects. This approach is preferable since it respects the inter-site differences that are present, whilst still combining the data from all of the sites within a sub-region, enabling us to make strong conclusions regarding that sub-region as a whole,which can then be compared. For this approach we choose a $N(\mu, \sigma^2)$ distribution, where μ and σ^2 are parameters within our model, which can therefore be specified by the data. We now fit a model to the "NW South" sub-region (containing 40 sites,one of which being "Heaton Park" which we looked at previously) where we allow the slope parameter μ_β to be a random effect, and hence have prior $N(\mu, \sigma^2)$ where μ and σ^2 each have vague $N(0, 1000^2)$ prior distributions. The code for this model is given in .6, and as with the scale parameter, we shall work with the log of the random effect standard deviation to avoid unnecessary rejection of proposed values and for extra precision in our conclusions we increase the number of iterations to 5000. Here we begin by studying the traceplots of the random effects parameters, shown in figure 5.1

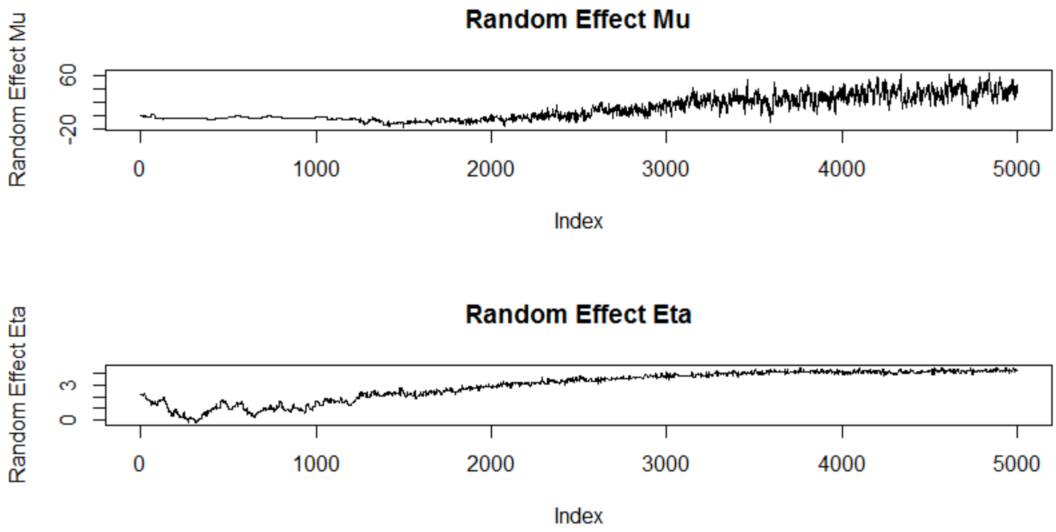

Figure 5.1: Traceplots of random effects parameters

In figure 5.1 we observe a phenomena which is common in large scale MCMC runs but as of yet has not occurred in the runs we have looked it, that of a burn-in period. This is essentially where the chain does converge to it's stationary distribution, it simply takes some time to arrive there, as can be seen in both plots in figure 5.1 the parameters do not occupy a single band of values, as we have seen in previous trace plots until approximately 3000 iterations in. This is not a problem and we simply correct it by removing the burn-in period (here the first 3000 values from each parameter) and replotting, as shown in figure 5.2
From figure 5.2 we now see that both traceplots demonstrate the type of pattern we are used to seeing in a converged chain, in as much as the values move frequently and occupy a single band of values, whilst the chain may not be mixing perfectly it is adequate for uses here. We can therefore use this latter section of the chain run to plot our histograms of the posterior densities, shown in figure 5.3

From figure 5.3 we can see that the histogram for μ, the mean of our random effects distribution has density which virtually is all positive, with a relatively even spread but a clear peak at around 25, which alone would imply that we have a positive temporal trend in this region, however when we look at the histogram for the standard deviaition $exp(\eta)$, we see that it has virtually all of its density between 40 and 90, implying that there is a

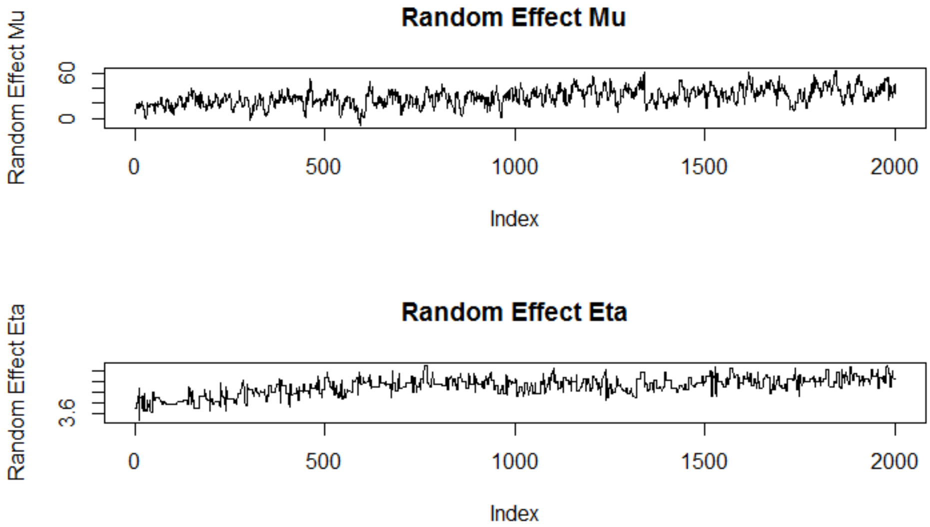

Figure 5.2: Traceplots with burn-in removed

fair amount of variability in the temporal trends in this region, and hence the standard deviation of our random effects distribution is greater than the mean. Therefore we do not have any real evidence that there is a clear, overall temporal trend in this region, which is already what we had observed at the "Heaton Park" site in the region.

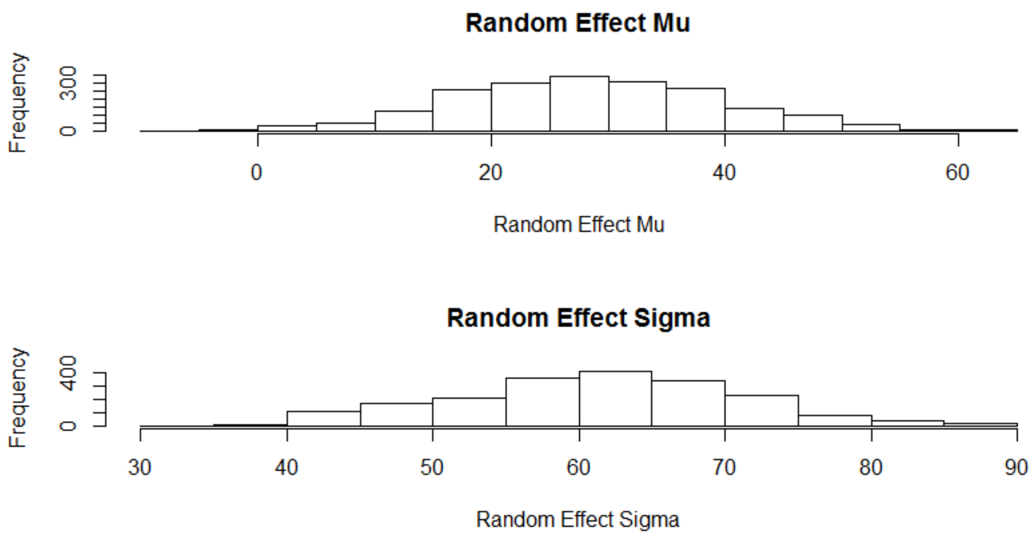

Figure 5.3: Histograms of posterior densities

Chapter 6

Conclusions

In this report we have investigated both a frequentist and Bayesian approach for modelling extreme rainfall by fitting GEV distributions to annual maxima, and showed how these approaches yield similar results for simplistic models, thanks to the vague prior assumptions we made, as well as showing how return levels can be estimated using both frequentist and Bayesian methods. From then we investigated the effects of assuming constant parameters within a localised area and showed how it dramatically reduced the variability of our posterior densities for the common parameters, thereby increasing the precision of our conclusions regarding them. Finally we investigated applying a random effects model for the temporal trend parameter in a localised region and showed how that can be used to draw conclusions regarding the presence, or lack thereof, of a temporal trend in these regions. We observed throughout the report that despite climate change events being reported, we did not see any clear evidence for a temporal trend in the location parameter of our GEV distribution at the sites we investigated, implying that the parameter may not be changing with time and the extreme rainfall events have not got significantly more severe over the course of the last 60 or so years.

A natural place to continue the study of this topic would be to apply a random effects term for each parameter in the model, since whilst we were reluctant to assume a constant parameter for the intercept and shape parameters across a sub-region, by using a random effects model we can still retain the individuality of each site, and any large differences in these parameters across a sub-region can be reflected by a large variance in the resulting random effects distribution. A mass application of this sort of model (regrettably not possible here due to the computing power/time required) could then be used in order to make some very useful inferences regarding flood estimation across the country. It may also be of interest to apply some Bayesian goodness of fit tests to discern just how well each model does in fact fit the data, which can be used as a method for testing the assumptions we have made when proposing each model (i.e. is it appropriate to assume a constant slope among sites for all sub-regions etc?). In addition, it would be desirable to achieve much longer runs of these chains (ideally of at least 10,000 runs) which we were not possible here due to computational restraints, which would provide even more firm conclusions with regards to our final posterior densities.

Chapter 7

Acknowledgements

I would like to thank the UK Meterological office and the UK Environment Agency for the data supplied for the CONVEX project funded by the NERC. I would also like to thank Chris Kilsby for supplying the original project proposal and Hayley Fowler and Stephen Blenkinsopp for setting up links to the data, and the UK Meteorological Office for providing the annual maximum data for Central and Eastern England on which I demonstrated some of my later models. Also my supervisor David Walshaw for the support and assistance he provided throughout the project, and for being a source of sanity when it was required. And finally I would like to thank the good people at Nescafe for getting me through those final few nights.

Bibliography

[1] Anis Ahmed. 100 dead and 250,000 stranded in bangladesh floods, June 2012.

[2] Stuart Coles. *An Introduction to Statistical of Extreme Values*. Springer Series in Statistics, 2001.

[3] Ronald Fisher and Leonard Tippett. On the estimation of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24(02).

[4] A.F. Jenkinson. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Proceedings of the Cambridge Philosophical Society*, 81(348).

[5] Conal Urquhart Matthew Weaver, Haroon Siddique. Uk storms: Met office issues red warning for damaging winds, February 2014.

[6] Richard Von Mises. La distribution de la plus grande de n valeurs. *Americal Mathematical Society*, 2:271–294, 1954.

[7] BBC News. In pictures: The story of hurricane katrina.

[8] CNN Staff. Typhoon haiyan death toll tops 6,000 in the philippines, December 2013.

[9] Kim Ann Zimmerman. Hurricane katrina: Facts, damage and aftermath, August 2012.

.1 Frequentist Code

```
install.packages(extRemes)
library(extRemes)
leap=c(rep(c(365,365,366,365),15),365,365,364)
leaps=c(1,1+cumsum(leap))
hourgev=function(x){
  print(x)
  data=read.table(x)[,7]
  totmiss=matrix(0,ncol=2,nrow=length(data)/24)
  maxmiss=matrix(0,ncol=2,nrow=63)
  max=vector()
  j=1
  for(i in 1:(length(data)/24)){
    totmiss[i,2]=sum(data[seq((24*i)-23,(24*i))]>=0)
  }
  for(i in 1:length(data)){
    if(data[i]<0)data[i]=0
  }
  for(i in 1:(length(data)/24)){
    totmiss[i,1]=sum(data[seq((24*i)-23,(24*i))])
  }
  for(i in 1:63){
    maxmiss[i,1]=max(totmiss[seq(leaps[i],leaps[i+1]-1),1])
    maxmiss[i,2]=sum(totmiss[seq(leaps[i],leaps[i+1]-1),2])
    if(maxmiss[i,2]>7300){
      max[j]=maxmiss[i,1]
      j=j+1
    }
  }
  if(length(max)>=5){
fit=gev.fit(max)
rlevel=return.level(fit,rperiods=c(10,50,200),make.plot=F)
  return(c(fit$conv,fit$mle,fit$se,rlevel$return.level,length(max)))
}
  else if(length(max)<5)
  {return(1:9)}
}

sites=list.files(pattern="*.*txt")
output=lapply(sites,hourgev)
```

.2 Frequentist Trend Code

```
yearfit=function(x){
  data=read.table(x)[,7]
  totmiss=matrix(0,ncol=2,nrow=length(data)/24)
  maxmiss=matrix(0,ncol=2,nrow=63)
  max=vector()
  years=vector()
  standyears=vector()
  year=1964
  j=1
  k=1
  for(i in 1:(length(data)/24)){
    totmiss[i,2]=sum(data[seq((24*i)-23,(24*i))]>=0)
  }
  for(i in 1:length(data)){
    if(data[i]<0)data[i]=0
  }
  for(i in 1:(length(data)/24)){
    totmiss[i,1]=sum(data[seq((24*i)-23,(24*i))])
  }
  for(i in 1:63){
    year=year+1
    maxmiss[i,1]=max(totmiss[seq(leaps[i],leaps[i+1]-1),1])
    maxmiss[i,2]=sum(totmiss[seq(leaps[i],leaps[i+1]-1),2])
    if(maxmiss[i,2]>7300){
      max[j]=maxmiss[i,1]
      j=j+1
      years[k]=year
      k=k+1
    }
  }
  for(i in 1:length(years)){
    standyears[i]=(years[i]-mean(years))/sd(years)
  }
  if(length(max)>=5){
  fit=gev.fit(max,ydat=matrix(standyears,ncol=1),mul=c(1))
  return(c(fit$conv,fit$mle,fit$se,length(max)))
  }
  else if(length(max)<5){
```

```
    return(1:10)
  }
}

sites=list.files(pattern="*.*txt")
output=lapply(sites, yearfit)
matout=do.call("rbind",output)
$
```

.3 Bayesian Code

```
gevll <- function(theta,dataset)
{
mu <- theta[1]
eta <- theta[2]
xi <- theta[3]
m <- 1+xi*(dataset-mu)/exp(eta)
if(min(min(m),exp(eta))<1e-5)return(-1e6)
loglik <- -length(dataset)*eta-sum(m^(-1/xi))-(1+1/xi)*sum(log(m))
return(loglik)
}
accept <- function(curr,prop,prior,j)
{
        accept <- FALSE
        if(runif(1)<exp(prop$lik-curr$lik)*dnorm(prop$theta[j],prior$m
        return(accept)
}
update <- function(curr,dataset,prior,j)
{
        prop <- curr
        prop$theta[j] <- rnorm(1,curr$theta[j],curr$err[j])
        prop$lik <- gevll(prop$theta,dataset)
        if(accept(curr,prop,prior,j))
        {
        curr <- prop
        curr$change[j] <- curr$change[j]+1
        }
        return(curr)
}
gevbayes <- function(n,dataset,start,err,mea,sdev)
{
```

```
        theta.mat <- matrix(0,nrow=n,ncol=3)
        curr <- list()
        curr$theta <- start
        curr$lik <- gevll(curr$theta,dataset)
        curr$change <- c(0,0,0)
        curr$err <- err
        prior <- list(mea=mea,sdev=sdev)
        for(i in 1:n)
        {
                for(j in 1:3) {curr <- update(curr,dataset,prior,j)}
                theta.mat[i,] <- curr$theta
        }
        feedback <- list(theta.mat,curr$change/n)
        return(feedback)
}
```

.4 Bayesian Trend Code

```
gevll <- function(theta,dataset)
{
  mua <- theta[1]
  mub <- theta[2]
  eta <- theta[3]
  xi <- theta[4]
  m <- 1+xi*(dataset-(mua+mub*time))/exp(eta)
  if(min(min(m),exp(eta))<1e-5)return(-1e6)
  loglik <- -length(dataset)*eta-sum(m^(-1/xi))-(1+1/xi)*sum(log(m))
  return(loglik)
}
gevbayes <- function(n,dataset,start,err,mea,sdev)
{
  theta.mat <- matrix(0,nrow=n,ncol=4)
  curr <- list()
  curr$theta <- start
  curr$lik <- gevll(curr$theta,dataset)
  curr$change <- c(0,0,0,0)
  curr$err <- err
  prior <- list(mea=mea,sdev=sdev)
  for(i in 1:n)
  {
    for(j in 1:4) {curr <- update(curr,dataset,prior,j)}
```

```
    theta.mat[i,] <- curr$theta
  }
  feedback  <- list(theta.mat,curr$change/n)
  return(feedback)
}
```

.5 Fixed Shape and Slope Code

```
gevllmult <- function(theta,dataset)
{
  mua <- theta[seq(1,(2*(length(dataset[1,])-1))+1,2)]
  eta <- theta[seq(2,2*length(dataset[1,]),2)]
  mub <- theta[(2*length(dataset[1,]))+1]
  xi <- theta[(2*length(dataset[1,]))+2]
  m=matrix(nrow=length(dataset[1,]),ncol=26)
  loglik=vector()
  for(i in 1:length(dataset[1,])){
    m[i,] <- 1+xi*(dataset[,i]-(mua[i]+mub*time))/exp(eta[i])
    if(min(min(m[i,]),exp(eta[i]))<1e-5)return(-1e6)
  }
  for(i in 1:length(dataset[1,])){
    loglik[i]=-length(dataset[,i])*eta[i]-sum(m[i,]^(-1/xi))-(1+1/xi)*
  }
  return(sum(loglik))
}
gevbayesmult <- function(n,dataset,start,err,mea,sdev)
{
  theta.mat <- matrix(0,nrow=n,ncol=(2*length(dataset[1,]))+2)
  curr <- list()
  curr$theta <- start
  curr$lik <- gevllmult(curr$theta,dataset)
  curr$change <- rep(0,(2*length(dataset[1,]))+2)
  curr$err <- err
  prior <- list(mea=mea,sdev=sdev)
  for(i in 1:n)
  {
    for(j in 1:((2*length(dataset[1,]))+2)) {curr <- updatemult(curr,d
    theta.mat[i,] <- curr$theta
  }
  feedback  <- list(theta.mat,curr$change/n)
  return(feedback)
```

```
}
```

.6 Random Effects Code

```
annmax=function(x){
  data=read.table(x)[,7]
  totmiss=matrix(0,ncol=2,nrow=length(data)/24)
  maxmiss=matrix(0,ncol=2,nrow=63)
  max=vector()
  years=vector()
  year=1949
  j=1
  k=1
  for(i in 1:(length(data)/24)){
    totmiss[i,2]=sum(data[seq((24*i)-23,(24*i))]>=0)
  }
  for(i in 1:length(data)){
    if(data[i]<0)data[i]=0
  }
  for(i in 1:(length(data)/24)){
    totmiss[i,1]=sum(data[seq((24*i)-23,(24*i))])
  }
  for(i in 1:63){
    year=year+1
    maxmiss[i,1]=max(totmiss[seq(leaps[i],leaps[i+1]-1),1])
    maxmiss[i,2]=sum(totmiss[seq(leaps[i],leaps[i+1]-1),2])
    if(maxmiss[i,2]>7300){
      max[j]=maxmiss[i,1]
      j=j+1
      years[k]=year
      k=k+1
    }
  }
  m=matrix(nrow=length(max),ncol=2)
  m[,1]=max
  m[,2]=years
  return(m)
}
gevllmult <- function(theta,dataset)
{
  s=length(dataset)
```

```
  mua <- theta[seq(1,3*(s-1)+1,3)]
  mub <- theta[seq(2,3*(s-1)+2,3)]
  eta <- theta[seq(3,3*s,3)]
  xi <- theta[3*s+1]
  m=vector()
  loglik=vector()
  for(i in 1:s){
    m <- 1+xi*(dataset[[i]][,1]-(mua[i]+mub[i]*((dataset[[i]][,2]-mean
     if(min(min(m),exp(eta[i]))<1e-5)return(-1e6)
     loglik[i]=-length(dataset[[i]][,1])*eta[i]-sum(m^(-1/xi))-(1+1/xi)
  }
  return(sum(loglik))
}
randll <- function(theta,dataset)
{
  mub <- theta[seq(2,(3*(length(dataset)-1))+2,3)]
  betamu <- theta[3*length(dataset)+2]
  betaeta <- theta[3*length(dataset)+3]
  randll=vector()
  for(i in 1:length(dataset)){
    randll[i]=log((1/(exp(betaeta)*sqrt(2*pi)))*exp(-(mub[i]-betamu)^2
  }
  return(sum(randll))
}
accept <- function(curr,prop,prior,j)
{
  accept <- FALSE
  if(dnorm(curr$theta[j],prior$mea[j],prior$sdev[j])==0) return(accept
  if(j<3*s+2){
   if(runif(1)<exp(prop$lik-curr$lik)*dnorm(prop$theta[j],prior$mea[j]
  }
  else{
  if(runif(1)<exp(prop$randlik-curr$randlik)*dnorm(prop$theta[j],prior$
  }
  return(accept)
}
updatemult <- function(curr,dataset,prior,j)
{
  prop <- curr
  prop$theta[j] <- rnorm(1,curr$theta[j],curr$err[j])
  prop$lik <- gevllmult(prop$theta,dataset)
```

```
  prop$randlik <- randll(prop$theta, dataset)
  if(accept(curr, prop, prior, j))
  {
    curr <- prop
    curr$change[j] <- curr$change[j]+1
  }
  return(curr)
}
gevbayesmult <- function(n, dataset, start, err, mea, sdev)
{
  s=length(dataset)
  theta.mat <- matrix(0, nrow=n, ncol=3*s+3)
  output <- matrix(0, nrow=n, ncol=5*s+3)
  curr <- list()
  curr$theta <- start
  curr$lik <- gevllmult(curr$theta[1:(3*s+1)], dataset)
  curr$randlik <- randll(curr$theta[(3*s+2):(3*s+3)], dataset)
  curr$change <- rep(0,(3*s)+3)
  curr$err <- err
  prior <- list(mea=mea, sdev=sdev)
  for(i in 1:n)
  {
    prior[[1]][seq(2,3*(s-1)+2,3)]=curr$theta[3*s+2]
    prior[[2]][seq(2,3*(s-1)+2,3)]=exp(curr$theta[3*s+3])
    for(j in 1:((3*s)+3)) {curr <- updatemult(curr, dataset, prior, j)}
    theta.mat[i,] <- curr$theta
  }
  feedback <- list(theta.mat, curr$change/n)
  return(feedback)
}
```