Analysis of Extreme Rainfall over Multiple Time Aggregations

Eleanor Kennedy

Supervisor:
Dr. Dave Walshaw

Abstract

Using extreme value theory to model environmental effects, specifically here rainfall across England and Wales, this paper will introduce the Generalised Extreme Value Theory which is then applied to 215 sites, split into six regions, to find likelihoods and maximum likelihood estimates for the parameters. Originally we had seven regions but due to the data undergoing quality control it has resulted in one region being discounted and a different number of sites being used for each region. The data that has been used is hourly rainfall counts but to take the investigation further, 24 hourly aggregations have been made thus we have daily data as well.

Due to similar geographical landscapes within each region, we have chosen to fit a model with a common shape parameter, ξ. However, this decision will be challenged with a second model, with varying ξ, where a Likelihood ratio test will be used to compare them.

After analysing the two models, the overall question that will be approached is:

"If we only have daily data available, can we say anything about what would happen for the hourly data?"

Contents

Chapter 1

Introduction

Exploring extreme weather is becoming increasingly more popular due to the practical effects it has on infrastructure, individuals and making inferences for the future. Different methods can be used to model these extremes in order to recognise trends and patterns over time or over different geographical regions. This paper will focus on rainfall levels in England and Wales split into six different regions. The data used will be introduced in Chapter 2 where scatterplots are seen to get a general feel for the data.

In order to explore the data in more detail Extreme Value Theory will be presented in Chapter 3, involving a brief history and the Extremal Types Theorem. From this the Generalized Extreme Value (GEV) Distribution is developed which is a fundamental result to this paper. Our chosen variable, rainfall, will then be modelled using the Block Maxima Approach, detailed in Chapter 3. To then examine the results from fitting our data, using the GEV distribution, MLE estimation and return level plots are introduced in Chapter 4.

Although the data is hourly rainfall we will be creating more datasets by aggregating these over 3, 6 12 and 24 hourly periods, thus creating 5 datasets for each site as opposed to just one. These will then be used to calculate parameter estimates and return level plots.

When fitting the data we will consider two cases: a common ξ for each region and a varying ξ or each site in all the regions. These two models will then be compared using the Generalised Likelihood Test to see which one is most appropriate for our data.

Suppose that we only have the 24 hourly totals for each site or region. The question proposed is then can we then say anything about the hourly data for the site or region? To answer this we will plot the parameter estimates, of the GEV distribution, that we found using the hourly data against those found using the 24 hourly data. This will hopefully result in a pattern which can then be modelled by a regression line overlaid on the plot. The same procedure will then repeated but instead of using parameter estimates we will plot the return levels using the hourly data against the return levels

found using the 24 hourly data.

We can also study depth-duration-frequency (DDF) curves which, for a given return period, describe the rainfall as a function of duration. It does this by plotting the return levels agains the return period, on a log scale, for each of the hourly aggregations. For one particular site we will plot these curves for each of the five time aggregations, giving us an idea of how the DDF curves change as we increase the hourly aggregations. We hope to see that the line plotted using the 24 hourly data will be highest and the line plotted using the houlry data will be the lowest. This will then prove useful to answer the question of can we say anything about the hourly data if we only have 24 hourly data.

Chapter 2

The Data

The data that will be studied in this paper are the hourly rainfall totals for 215 sites across England and Wales from 1949 to 2011. England and Wales have been split up into seven regions namely North East, North West, Midlands, Wales, Anglian, South West and South East, seen in Figure 2.1.

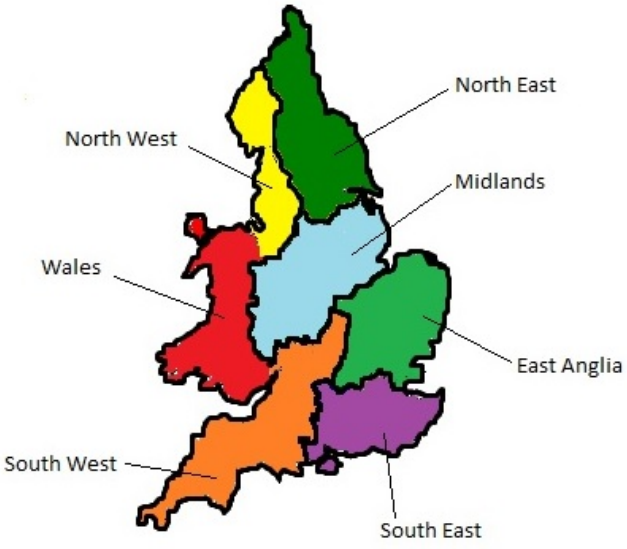

Figure 2.1: Regions of England and Wales

A quality control constraint has then been considered so that any site that has less than 85% of data for each year present and less than 85% of years present will be excluded resulting in the whole of the South West region being excluded from the study. In additon to this, within each region some sites have been discounted due to more than 15% of each site's data being missing. Once we have found the sites that have more

than 85% of the data present we then need to look at the percentage of data present for each year within that site. Again, if more than 15% of a year is missing then this year will be excluded. This results in varying amounts of data used for each site.

Table 2.1 below shows how many sites we have used in each region.

Region	Sites Used
Anglian	13
Midlands	34
North East	27
North West	25
South East	87
Wales	29

Table 2.1: The number of sites used in each region.

Although the data was collected hourly; 3, 6, 12, and 24 hourly aggregations can also be explored leading to a greater understanding of the data. Looking at the general trends and patterns over the regions is useful but as each region has a different number of sites we will just pick one from each region to explore. The figures below show the annual maximum for the five different aggregations explained above for the years there is data present. One site for each region has been picked as a representative for that region.

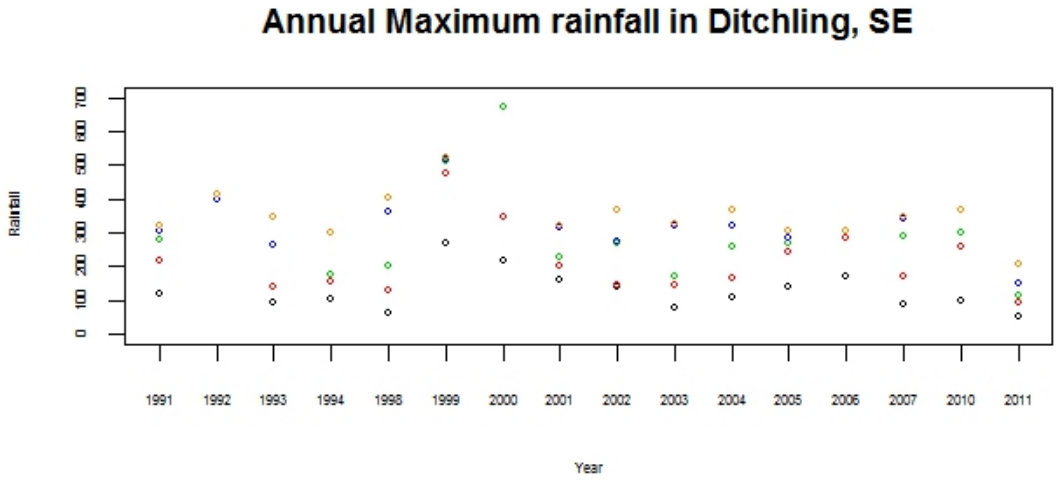

Figure 2.2: Annual maximums for Ditchling, SE. Black represents hourly data, red represents 3 hourly aggregations, green is 6 hour aggregations and blue and orange are 12 and 24 aggregations respetively

Looking in general at the plots we see the annual maximum of rainfall increases as we aggregate over longer time periods (represented by the different colours), as must be the

case. In addition, for each site the amount of rain falling in each year is similar with some peaks in some of the regions. For example, looking at Ditchling in South East England, in Figure 2.2 we a see a peak in 1999 and for Birchmoor in East Anglia, in Figure 2.3 there is a peak in 1992.

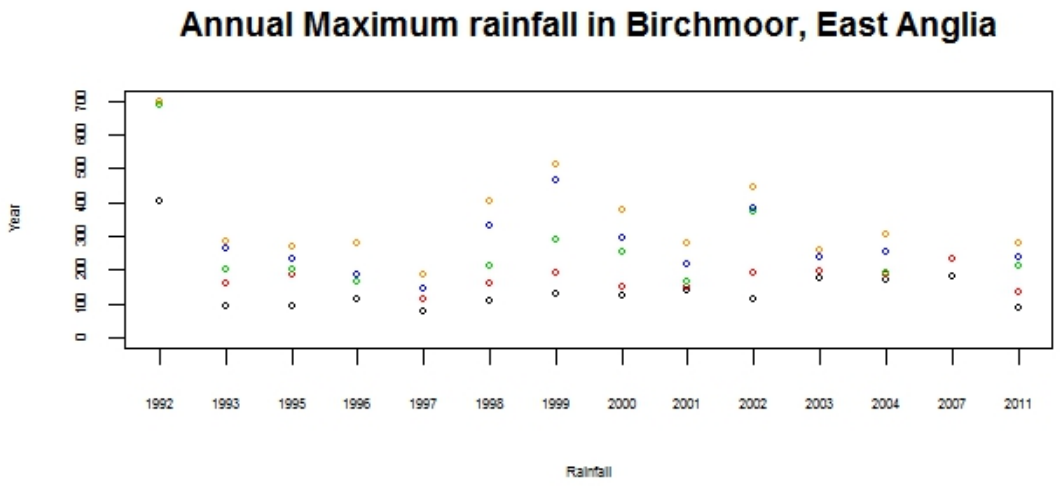

Figure 2.3: Annual maximums for Birchmoor, Anglian. Black represents hourly data, red represents 3 hourly aggregations, green is 6 hour aggregations and blue and orange are 12 and 24 aggregations respetively

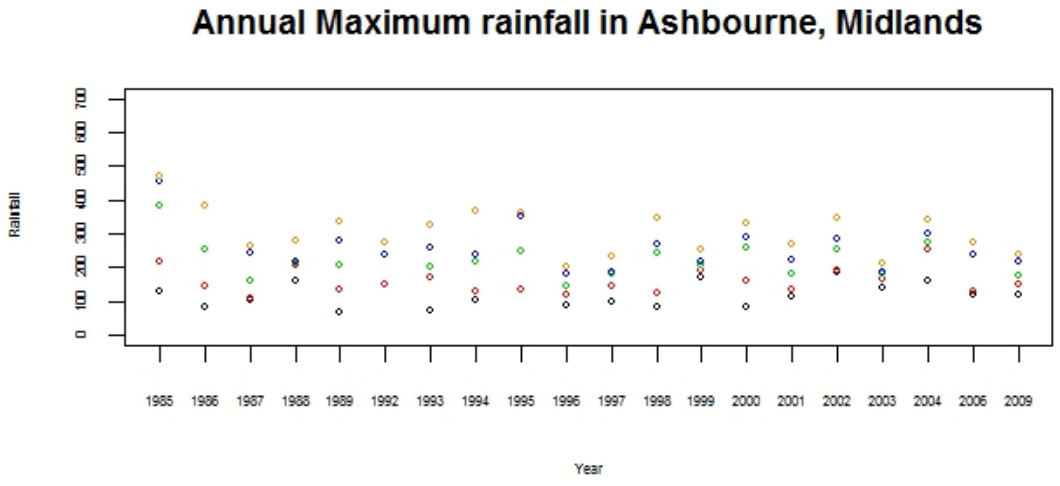

Figure 2.4: Annual maximums for Ashbourne, Midlands. Black represents hourly data, red represents 3 hourly aggregations, green is 6 hour aggregations and blue and orange are 12 and 24 aggregations respetively

Taking a step back we can see that some sites, as a whole, have a greater amount of rainfall year on year. Comparing Ashbourne in the Midlands, in Figure 2.4 above, to Jesmond in North East England, in Figure 2.5, below, we can see that the annual maximum rainfall, on average, is much greater in Jesmond.

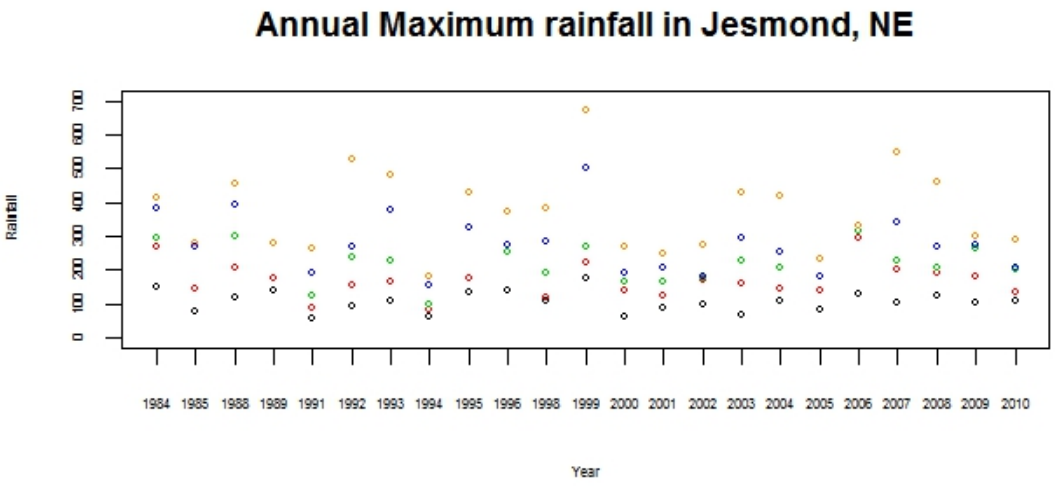

Figure 2.5: Annual maximums for Jesmond, NE. Black represents hourly data, red represents 3 hourly aggregations, green is 6 hour aggregations and blue and orange are 12 and 24 aggregations respetively

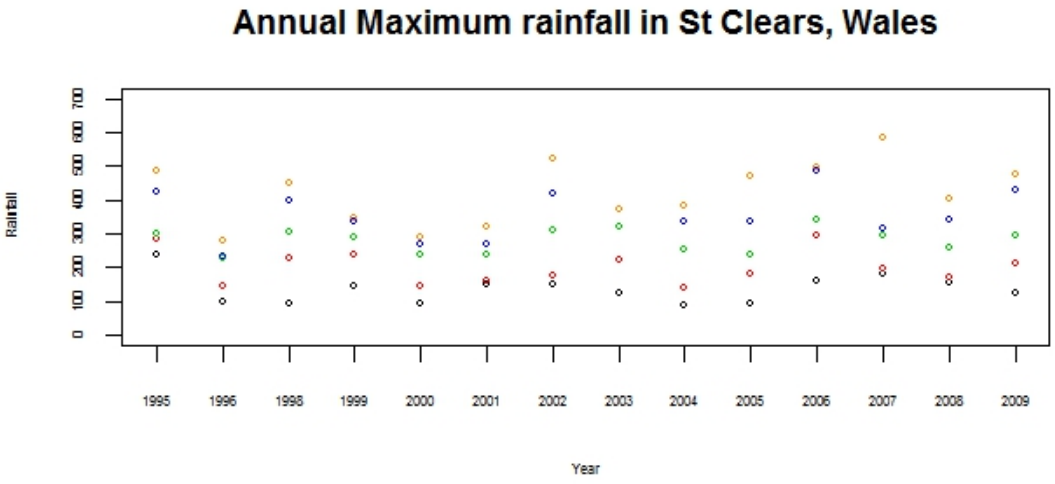

Figure 2.6: Annual maximums for St Clears, Wales. Black represents hourly data, red represents 3 hourly aggregations, green is 6 hour aggregations and blue and orange are 12 and 24 aggregations respetively

In Figure 2.6 above we can see that St Clears in Wales has annual maximums somewhere between the Midlands and the North East, remaining steady with some high peaks in 2000 and 2007. Haighton, in the North West is similar with an obvious large amount of rainfall in 1995, seen in Figure 2.7

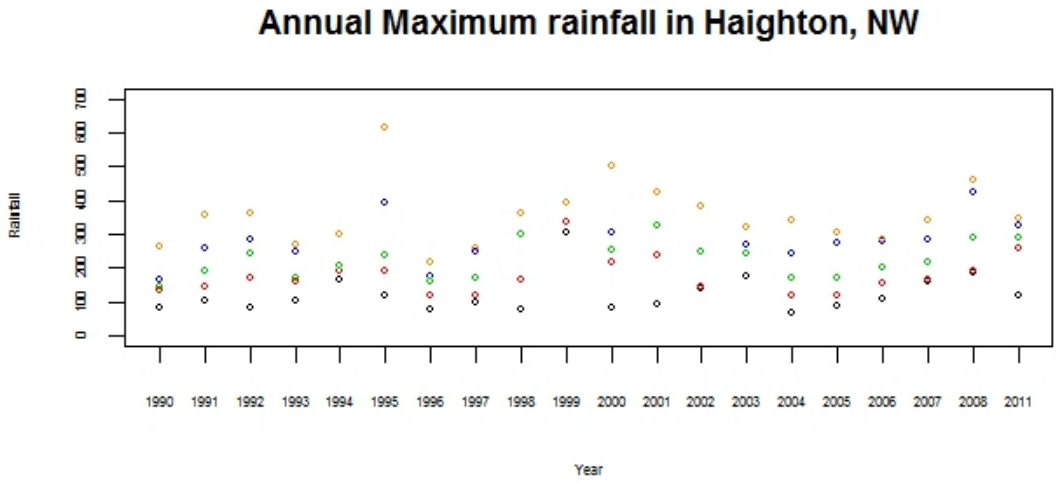

Figure 2.7: Annual maximums for Haighton, NW. Black represents hourly data, red represents 3 hourly aggregations, green is 6 hour aggregations and blue and orange are 12 and 24 aggregations respetively

Although these plots give us insight into some patterns and trends that we think might be happening regionally we need to look at all the sites in each region to get the full picture. Later in this report we will apply extreme value theory to analyse all the data fully.

Chapter 3

Extreme Value Theory

3.1 A brief History

Although it was only first published in Gumbel (1958), who was applying the theory to engineering, Extreme Value Theory (EVT) has been traced as far back as 1709 where mathematicians were using it to find the longest survivor in a group of people. Further investigation was then carried out by Leonard Tippett who, whilst he was working for the British Cotton Industry Research Association, realised that the weakest fibres were responsible for the overall strength of a peice of cotton. This idea was then developed, and with the help of E.J. Gumbel and R. Fisher, he obtained three extreme value distributions. In spite of the fact that these three men introduced these ideas it was only proved later in Gnedenko (1943).

Throughout the 1970's the theory was developed to form the basis of the statistical models we are going to use, generalised by L. de Haan and J Picklands. The theory and models are still being investigated to date causing Extreme Value Theory to be more well known world-wide with an increase of practical applications.

3.2 The Extremal Types Theorem

In order to use the GEV distribution, mentioned in Chapter 1, it is useful to discuss the central result in EVT in more detail. It is analogous to the Central Limit Theorem (CLT) however here we will work with the maximum values as opposed to the mean values. It enables us to estimate the probability of extreme events given a previously observed ordered sample. For example, Extreme Value Theory might be used by engineers to estimate the amount of rainfall we expect to see every one hundred years therefore enabling them to decide what capacity is needed for a drainage system.

More formally, suppose $X_1, X_2, ...$ is a sequence of random variables that are independent and identically distributed and take $M_n = \max(x_1, x_2, \dots)$. The EVT looks at the limiting distribution of M_n as we increase n. The question then follows of what distributions can be considerd for M_n as $n \to \infty$?

The limiting distribution of M_n will always converge to a single point, in other words the distribution is degenerate. As stated above this is analogous to the Central Limit Theorem where the sample mean $\bar{X}$ converges to the population mean μ. In the CLT this is prevented by normalising the random variable so that

$$\frac{\bar{X} - b_n}{a_n} \xrightarrow{D} N(0,1)$$

where σ is the population standard deviation and n is the population sample size and $b_n = \mu$ and $a_n = \sigma/\sqrt{n}$.

Similarly, this can be applied to our maximums M_n giving rise to the Extremal Types Theorem.

Theorem 1 (The Extreme Value Theory)**.** *Suppose we have a sequence of constants $a_n > 0$ and b_n such that*

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

Let G be a non-degenerate distribution function that belongs to one of the following families:

$$1: \quad G(z) = exp\Big\{ - exp\Big[- \Big(\frac{z-\beta}{\gamma}\Big)\Big]\Big\}, \quad -\infty < z < \infty; \tag{3.1}$$

$$2: \quad G(z) = exp\Big\{ - \Big(\frac{z-\beta}{\gamma}\Big)^{-\alpha}\Big\}, \quad z > \beta; \quad [G(z) = 0, z \leq \beta]; \tag{3.2}$$

$$3: \quad G(z) = exp\Big\{ - \Big[- \Big(\Big(\frac{z-\beta}{\gamma}\Big)^{\alpha}\Big]\Big\}, \quad z > \beta; \quad [G(z) = 1, z \leq \beta]; \tag{3.3}$$

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

The three types of distribution, given above, are known as Type 1, the Gumbel Distribution; Type 2, the Frèchet distribution and Type 3, the Weibull distribution.

Unlike the Weibull distribution, both Gumbel and Frèchet's upper-endpoints tend to ∞ and so the distribution G is unbounded. It cannot always be guaranteed that there is an existence of a limiting distribution however when there is we know that

$$\frac{M_n - b_n}{a_n} \xrightarrow{D} G$$

where G is one of the above distributions. The question is now, which distribution do we use?

3.3 The Generalised Extreme Value Distribution

Deciding which of the three distributions to use can be problematic and so we consider a reparameterisation of all three, known as the Generalised Extreme Value Distribution.

The GEV has cumulative distribution function

$$G(z) \;=\; \exp\Big\{ \;-\; \Big[\, 1+\xi\,\Big(\,\Big(\frac{z-\mu}{\sigma}\Big)^{-1/\xi}\Big\}, \tag{3.4}$$

where $z \;:\; [1+\xi(z-\mu)/\sigma] > 0$ and $\mu,\ \sigma \;>\; 0$ and ξ are the location, scale and shape parameters of the distribution. This can be seen in Jenkinson (1955) and Von Mises (1956).

For different values of ξ the heaviness of the tail changes. More specifically, for $\xi > 0$ the distribution has a lower end point but no finite upper end point and the c.d.f is only valid for $x > \mu - \sigma/\xi$. For $\xi < 0$ the distribution has an upper end point but no finite lower end point and the c.d.f. is only valid for $x < \mu + \sigma/\xi$. However, if $\xi = 0$ the c.d.f. above is formally undefined and so we take the limit of the c.d.f. as $\xi \to 0$, giving

$$G(z) \;=\; \exp\Big\{ -\exp\Big[\;-\; \Big(\frac{z-\mu}{\sigma}\Big)\Big]\Big\}, \tag{3.5}$$

defined for all z. The location and scale parameters affect both the mean and variance of the disribution.

In section 3 we introduced the constants a_n and b_n and we know that

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} \mathcal{G}(\mu,\sigma,\xi), \quad \text{as } n \to \infty,$$

where $\mathcal{G}$ is the c.d.f. given in Equation 3.5. This can be simplified to

$$M_n \xrightarrow{d} \mathcal{G}(\mu^*,\sigma^*,\xi), \quad \text{as } n \to \infty$$

where a_n and b_n have been included into μ^* and σ^*. We do not need to worry about these normalisation constants as the GEV parameters are estimated anyway and we can just fit the GEV to the set of maximums.

The GEV distribution can be applied for our data provided the conditions for it to apply hold in our case. However, the conditions are rather weak. Provided the maxima are selected from a continuous non-degenerate distribution with finite mean and variance, then we reasonably expect the GEV to be the correct limit.

Chapter 4

Fitting the GEV distribution

To estimate the parameters in the GEV several different methods can be used. One example is the method of moments which calculates the k^{th} moments $E(X^k)$ and then equates them to sample values $\sum_{i=1}^{n} \frac{x_i^k}{m}$. The equations are then solved to give the parameter estimates. Another method is to use maximum likelihood estimation. The latter is what this report will be focusing on.

4.1 Maximum Likelihood Estimation

The likelihood of the data with respect to μ, σ and ξ is

$$L(\mu, \sigma, \xi|y) \ = \ \prod_{i=1}^{n} f(y|\mu, \sigma, \xi),$$

and so the log likelihood is

$$l(\mu, \sigma, \xi|y) \ = \ \log\{L(\mu, \sigma, \xi|y)\}.$$

After differentiating this function with respect to μ, σ and ξ and setting each of the derivatives to zero, the equations can then be rearranged to give the maximum likelihood estimates $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}$.

In this report we will investigate a simplification to the model that can be made by fixing the shape parameters within each of the six regions studied. For example, under this assummption, all the of the sites in the North West will have the same ξ value but different location and scale parameters. This allows for local differences in mean and variance, but imposes the same tail behaviour across the region, corresponding to similar geographical conditions.

4.2 Return Level Curves

Although it is useful to find the real values of μ, σ and ξ people are more likely to be interested in predicting the probability of an event happening in a period of time, promoting the need for return level curves.

The r-year return level, $z(r)$, is defined as a value that is expected to be exceeded, on average, once every r years. These are found by setting the GEV distribution function (Equation 4) equal to $1 - 1/r$ and solving for $x = \hat{z}_r$.

For example, after fitting the GEV to a set of annual maxima we obtain estimates for $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}$. The 50-year return level can be calculated as follows. Firstly the following probability statement can be made:

$$Pr(\text{annual maximum} \leq z_{50}) \;=\; 1 - \frac{1}{50} = 0.98. \tag{4.1}$$

As $Pr(\text{annual maximum} \leq z_{50}) = G(z_{50}; \hat{\mu}, \hat{\sigma}, \hat{\xi})$ we can now express Equation 3.4 as

$$1 - \exp\Big\{ \;-\; \Big[\, 1 + \hat{\xi}\,\Big(\,\Big(\frac{\hat{z}_{50} - \hat{\mu}}{\hat{\sigma}}\Big)^{-1/\hat{\xi}}\Big\} \;=\; \frac{1}{50}.$$

By rearranging this equation an estimate of the 50^{th}-year return level is found to be

$$\hat{z}_{50} \;=\; \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}\Big[\Big(-\log(0.98)\Big)^{-\hat{\xi}} - 1\Big].$$

This is then generalised to give an estimate of the r-year return level as

$$\hat{z}_r \;=\; \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}\Big[\Big(-\log(1 - r^{-1})\Big)^{-\hat{\xi}} - 1\Big]. \tag{4.2}$$

Once the return levels have been found they are then plotted against the return period. The return period is expressed on a log scale enabling the return levels to be visualised easily for a range of return periods.

Chapter 5

Maximum Likelihoods fitting for the Data

Fitting the GEV distribution to each site within each of the six regions yeilds an MLE for the parameters, μ, σ and ξ, at each time aggregation i.e using hourly data, 3, 6, 9 and 12 hourly aggregations and daily data (24 hourly aggregations). Here it is not possible to look at the values for these MLEs as there are too many, instead it is more useful to look at plots of them comparing each region.

To obtain these maximum likelihood estimates for μ, σ and ξ we use the statistical package R [R Core Team, 2013] . A programme is written, seen in the Appendix, which iterates the negative log likelihood with varying values of μ and σ and a common ξ until it eventually obtains the minimum. The parameters are then displayed along with the actual minimum point. Later we will consider the log likelihoods using varying values of ξ and then compare them using the Generalised Likelihood Ratio Test to see if the assumption of common ξ is reasonable.

To get an overall feel for the parameter estimates calculated we can plot them regionally. Here we are not so much interested in the actual values but the patterns and trends that are apparent. Following on from this we can then look at the relationship between the parameter estimates calculated using the hourly aggregations and those calculated using the 24 hourly aggregations. A regression line can then be fitted to clearly display this relationship.

5.1 Regional Parameter Estimates

For each region there are three scatterplots showing the MLE's for μ, σ and ξ against each time aggregation where each site is represented by a point. Due to varying numbers of sites measured for each region μ, and σ will have a different number of points. On the other hand, the plots for ξ will only have 5 points on each plot, one for each time

aggregation. This is because we have assumed a common shape parameter for each site due to the similar geographical landscape of each region, the only variation here is the time aggregations.

First looking at the north of England, Figure 5.1 shows the MLEs for the North West on the top row and the North East on the bottom row. Both plots for $\hat{\mu}$ show a steady increase as the time aggregations are increased. We can also see that the dispersion of points is greater the more hours you aggregate.

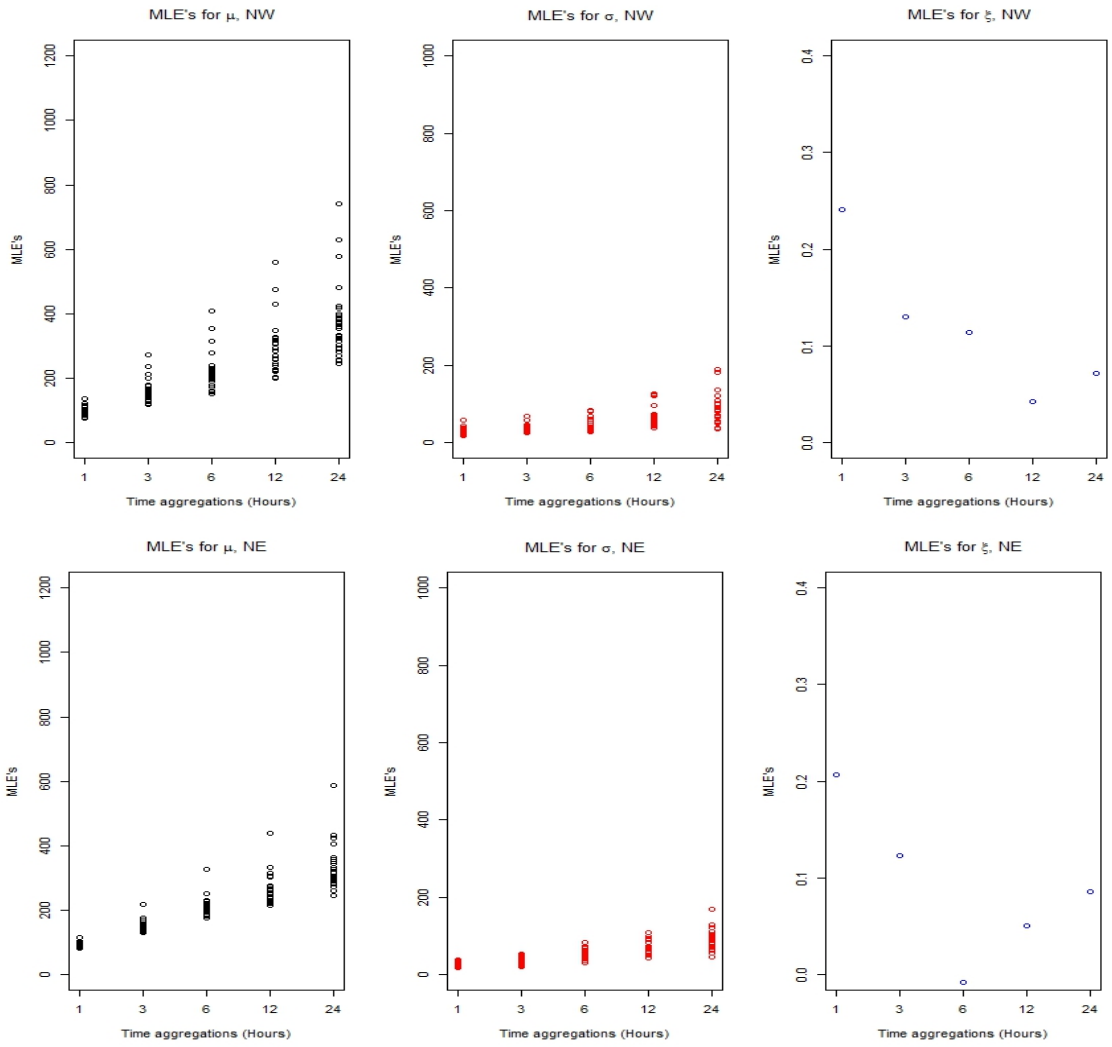

Figure 5.1: MLEs for the GEV parameters of the North West and North East regions

Similarly the values for $\hat{\sigma}$ are increasing as we move along the x-axis, again with increasing dispersion. In contrast to $\hat{\mu}$, the $\hat{\sigma}$ values are smaller but as the scales are different for both $\hat{\mu}$ and $\hat{\sigma}$ we cannot make a direct comparison from the plots. The scales have been chosen in such a way that we can compare the $\hat{\mu}$ values for each region against

each other, the $\hat{\sigma}$ values against each other and the $\hat{\xi}$ values against each other rather than the $\hat{\sigma}$ values against the $\hat{\sigma}$ values regionally.

The shape parameter, on the other hand, is expected to decrease as the time aggregations increase [Katz, 2011].Looking at the North East in particular, we can see a dip when using the 6 hourly aggregations before increasing again at 12 and 24 hourly aggregations. Although we see variation throughout the time aggregations we can confirm that in both plots, generally, the values for $\hat{\xi}$, the blue points, are decreasing.

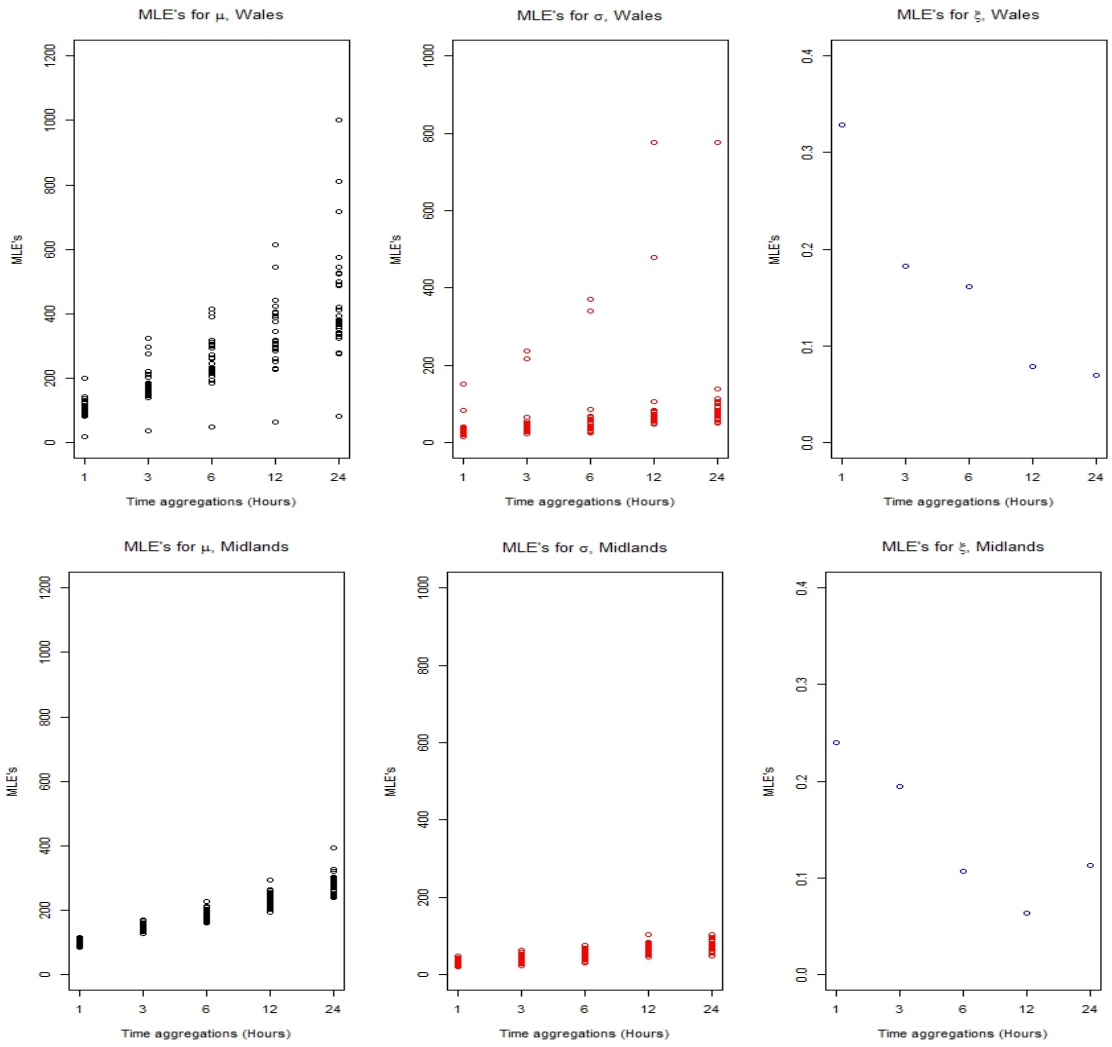

Figure 5.2: MLEs for the GEV parameters of Wales and the Midlands

Moving further south geographically, Figure 5.2 shows the MLEs of the parameters for Wales and the Midlands. Looking at the plots for Wales, although one site's μ estimates are staying steady as we increase the time aggregations the others are increasing as seen in both the North West and East. On the other hand, the majority of the values for $\hat{\sigma}$

are invariant with the aggregations apart from two of the sites represented by the two higher points seen at each time point. In further work, we may choose to investigate these two sites and see if we could explain the large values for $\hat{\sigma}$. Again, as with the two most northern sites, the values for $\hat{\xi}$ are fluctuating between 0 and 0.4 with what seems like a decrease as the time aggregations are increased.

The Midlands show a uniform increase in $\hat{\mu}$ values however, in comparison to the other three regions we have already looked at, the highest value is only 394.29. This low value of parameter estimate continues into the plot for $\hat{\sigma}$ as although the values are increasing they are increasing at a relatively low rate. Again, the MLEs for ξ are between 0 and 0.4 with a decreasing pattern.

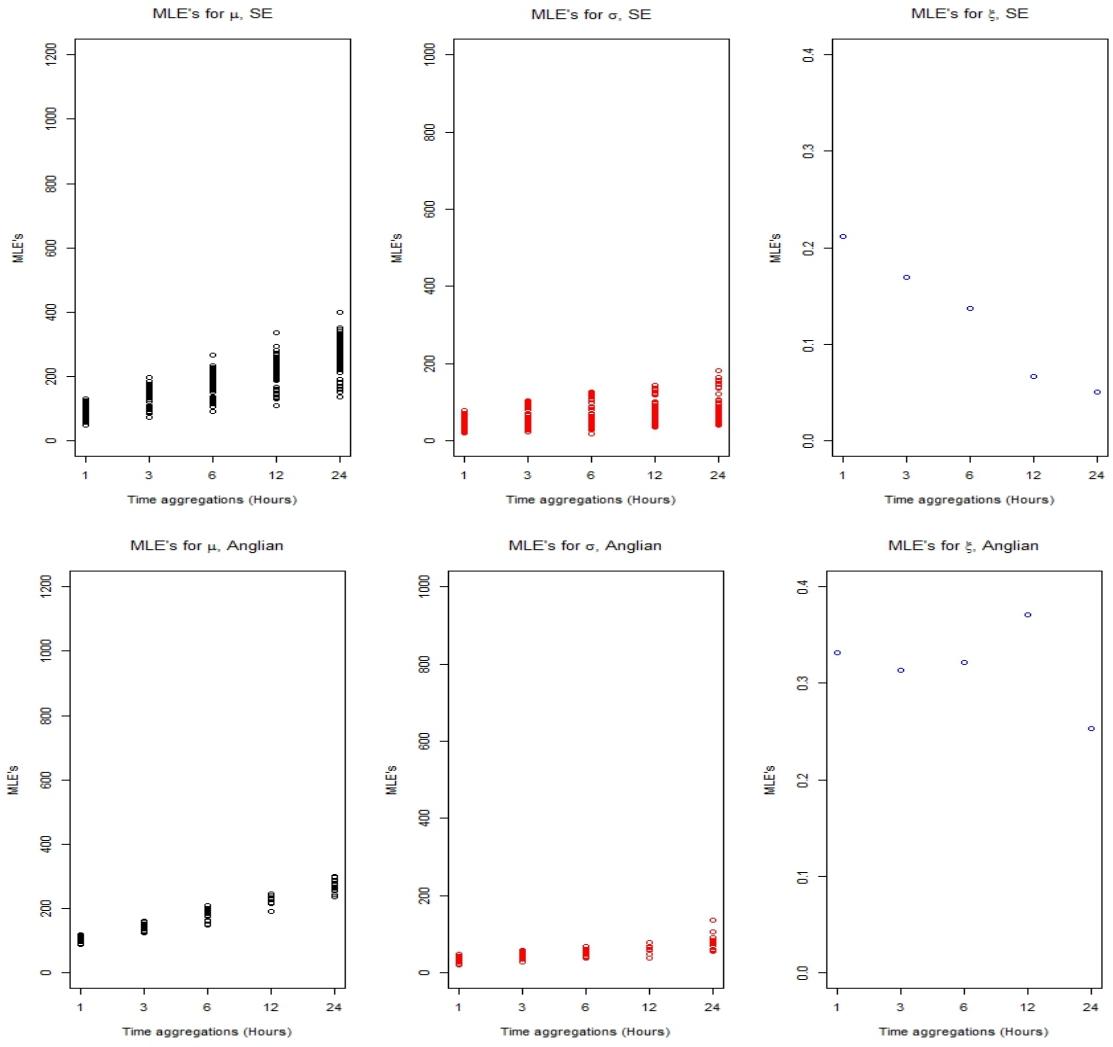

Figure 5.3: MLEs for the GEV parameters of the South East and Anglian regions.

In Figure 5.3, above, the increase in the values for $\hat{\mu}$ and $\hat{\sigma}$ as we increase the time

aggregations for the South East region are similar to those seen in the two northern regions but with less dispersion of points. This may be due to the South East having the highest number of sites used in this data and so the points just appear closer together because there is more. The values for $\hat{\xi}$ are following the same pattern that those for the Midlands did with a slight decrease but maintaining between 0 and 0.4.

The Anglian region, seen on the bottom row, has increasing $\hat{\mu}$ and $\hat{\sigma}$ parameters however they are increasing at the lowest rate in comparison with the other five sites. The values for $\hat{\xi}$ are oscillating between 0.2 and 0.4 but in comparison to the other sites we don't see as much of a decrease from the hourly to the 24 hourly aggregations. This may be due to the Anglian region have significantly less sites than the others, thus the expected pattern is not so obvious.

In conclusion we have seen that for $\hat{\mu}$ and $\hat{\sigma}$ the values increase as we increase the time aggregations along with a greater dispersion of points. Both the Anglian region and the Midlands are increasing at the lowest rate and Wales and the North West are increasing with the highest rate.

For all regions the parameter estimates for ξ appear to decrease as we increase the time aggregations. To investigate further, the MLEs for each region have been plotted on the same graph which can be seen in Figure 5.4. From time aggregations 1 to 6 the MLEs look steady with no sudden change however, at time aggregation 12 the Anglian region seems to increase and the other 5 regions seem to decrease. The opposite seems to occur when we use the 24 hourly aggregations i.e. daily data. Although we have an increase from the 12 hourly aggregations to the 24 hourly aggregations, overall the values for $\hat{\xi}$ are thus confirming what was expected.

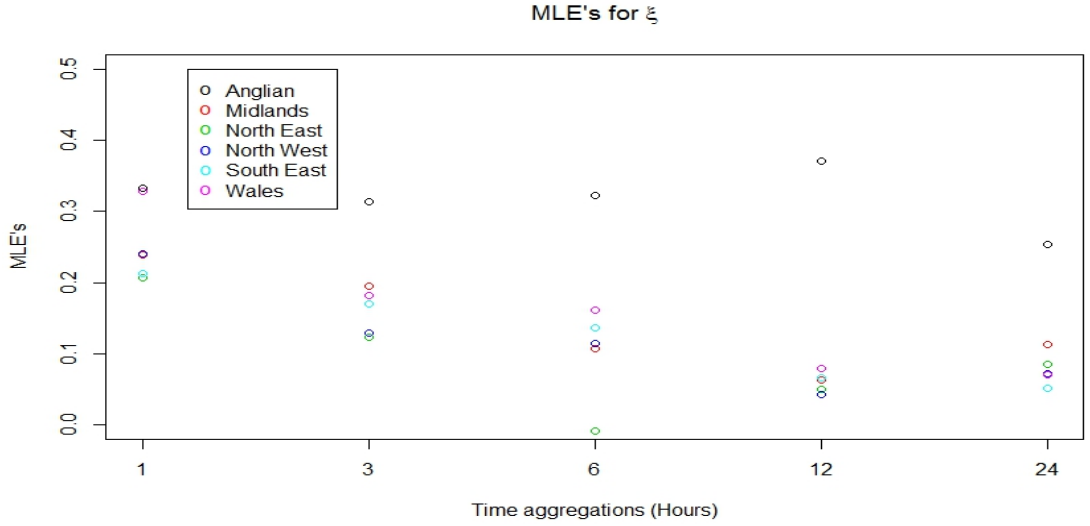

Figure 5.4: MLEs for the GEV parameter ξ for all regions.

Just taking the hourly and 24 hourly aggregations, i.e. daily, the values for these MLEs, along with the standard errors, can be seen in Table 5.1 below. If we look at each region seperately we can see that the MLE for ξ is always lower when using the daily data rather than the hourly data therefore supporting our conclusion that the $\hat{\xi}$ values do depend on the time aggregation.

Looking at the raw values for the estimates is fine but we do not know anything about the variability of these values and so it is useful to find the standard errors (s.e's) for each MLE. Using the nlm function in R we obtain the negative second order partial derivatives of the function we are minimising, thus giving us the observed information matrix needed to calculate the standard errors.

This matrix, evaluated at $\hat{\mu}$, $\hat{\sigma}$ and $\hat{\xi}$ is then inverted to give the Variance-Covariance matrix where the variances are the diagonal elements. Finally, to obtain the standard errors these values are square rooted. This can then be repeated for each region with the values seen below in Table 5.1.

Region	$\hat{\xi}$ (Hourly)	s.e.($\hat{\xi}$)	$\hat{\xi}$ (Daily)	s.e.($\hat{\xi}$)
Anglian	0.3321	0.0784	0.2528	0.0745
Midlands	0.2396	0.0349	0.1133	0.0293
North East	0.2069	0.0409	0.0853	0.0394
North West	0.2406	0.0544	0.0713	0.0380
South East	0.2121	0.0233	0.0505	0.0200
Wales	0.3288	0.0476	0.0697	0.0435

Table 5.1: The MLEs for ξ using the hourly data and the 24 hourly aggregations along with the standard errors.

The standard errors are important here to see if our estimates of ξ are significant or not. To test this we need to calculate the 95% confidence intervals for each $\hat{\xi}$ by using,

$$\left(\hat{\xi} \; - \; 1.96 \; \times \; s.e.(\hat{\xi}), \; \hat{\xi} \; + \; 1.96 \; \times \; s.e.(\hat{\xi})\right).$$

The confidence intervals can be seen in Table 5.2. The general rule is to check to see whether the $\hat{\xi}$ using the hourly data comes within the confidence interval for $\hat{\xi}$ using the daily data and vice versa. Here we can see that for the Anglian region both $\hat{\xi}$ values are contained within the opposite confidence interval, hence it is not significant. On the other hand, the other 5 regions are all significant with none of the $\hat{\xi}$ values coming in the confidence region for the opposing time aggregation.

Region	$\hat{\xi}$ CI (Hourly)	$\hat{\xi}$ CI (Daily)
Anglian	(0.1784, 0.4858)	(0.1068, 0.3988)
Midlands	(0.1712, 0.3080)	(0.0553, 0.1713)
North East	(0.1267, 0.2871)	(0.0081, 0.1625)
North West	(0.1340. 0.3472)	(-0.0032, 0.1458)
South East	(0.1664. 0.2578)	(0.0113, 0.0897)
Wales	(0.2355, 0.4221)	(-0.0156, 0.1550)

Table 5.2: The 95% confidence intervals for $\hat{\xi}$ using hourly and daily data

5.2 Regression of Parameter Estimates

After looking at the general pattern for the parameter estimates we can now focus on the relationship between the hourly and the 24 hourly aggregations. The following plots are the parameter estimates for μ and σ based on the hourly data against the estimates using the 24 hourly data, where each point represents a site in that given region. As we have assumed a common ξ for each region we will not be looking at plots for this parameter.

From the data, linear regression can then be used to model the dependent variable, which here is the parameter estimate using the hourly aggregations, and the explanatory variable, here the parameter estimates found using the 24 hourly aggregations. We have that the 24 hourly aggregations are a function of the hourly data but we are trying to make inferences on the inverse problem i.e. if we only have the 24 hourly data what can we say about the 1 hour aggregations. This line is found by fitting a linear model to the data using the statistical programme R with the commands and output seen in the Appendix.

Firstly we can see the MLEs for μ and σ for the North West region on the top row of Figure 5.5 below. For μ the estimates from the hourly and 24 hourly aggregations have high positive correlation, illustrated by the red linear regression line. The estimates for σ, on the other hand, are only slightly positively correlated with no real trend. This will make it difficult if we only had the 24 hourly totals but wanted to predict what the parameter estimates would be if we had the hourly totals.

Similarly, $\hat{\mu}$ and $\hat{\sigma}$ for the North East region are both positively correlated but with the estimates for μ having greater correlation than the estimates for σ. Again making it difficult to make inferences from just the estimates found by using the 24 hourly totals.

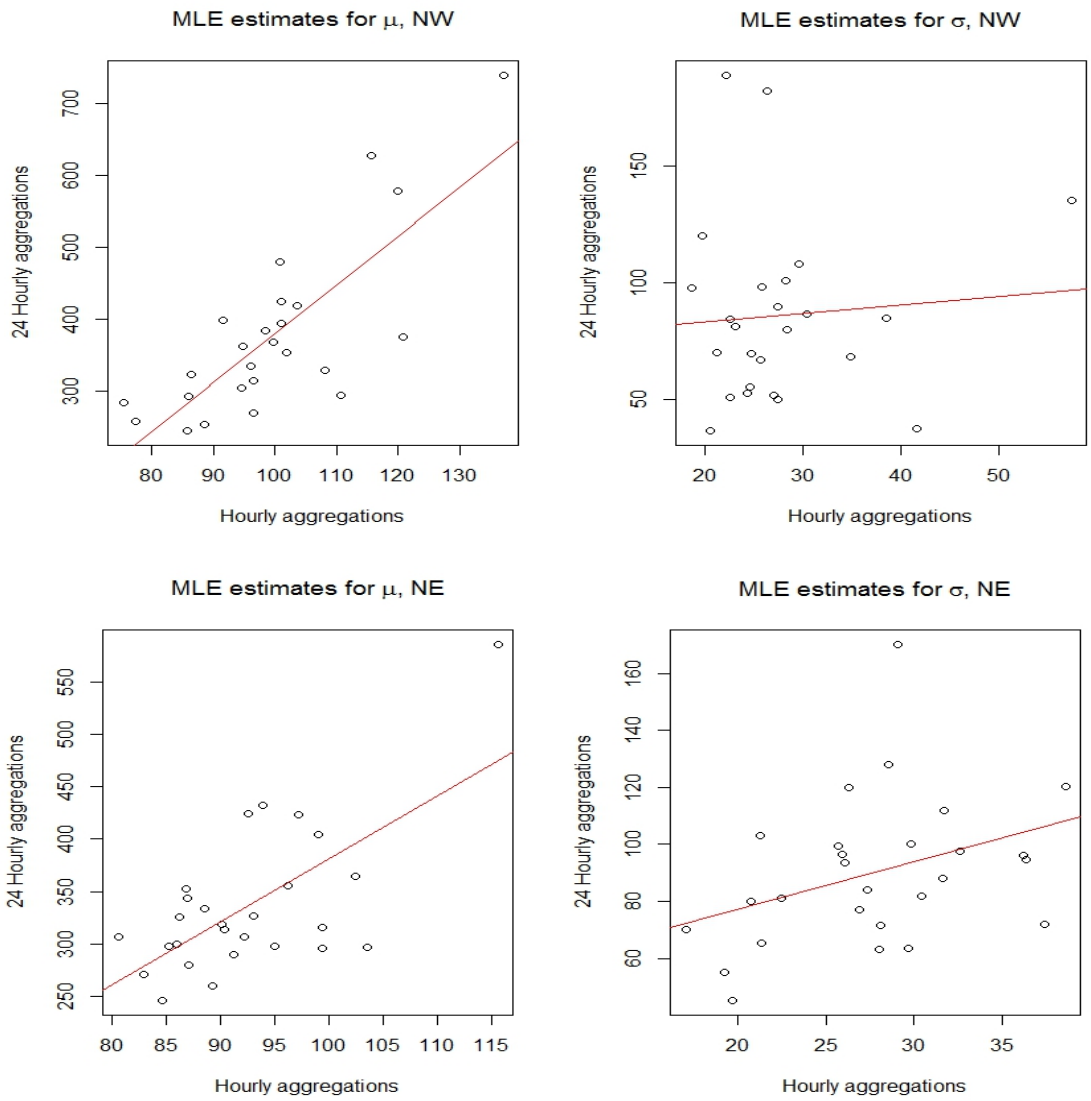

Figure 5.5: MLEs for the GEV parameters using the hourly aggregations plotted against those using the 24 hourly aggregations for the North West and North East region.

Similar plots can be seen for Wales and the Midlands in Figure 5.6 below. Firstly, the $\hat{\mu}$ values found using the hourly aggreations are positively correlated with those using the 24 hourly aggregations with the majority of the sites clustered in the centre of the plot. For $\hat{\sigma}$ the main cluster of points is in the bottom left corner with seemingly no correlation, contrasted with the regression line showing high positive correlation. Care needs to be taken when using this regression line as it is only drawn like this due to the two outliers in the top right corner.

The Midlands, on the other hand, have no obvious pattern to either plot with no major clustering of points. There is slight positive correlation in both plots with the estimates

for μ having marginally higher correlation than the estimates for σ. This again would make it difficult to estimate the parameters using hourly data if we only have 24 hourly data.

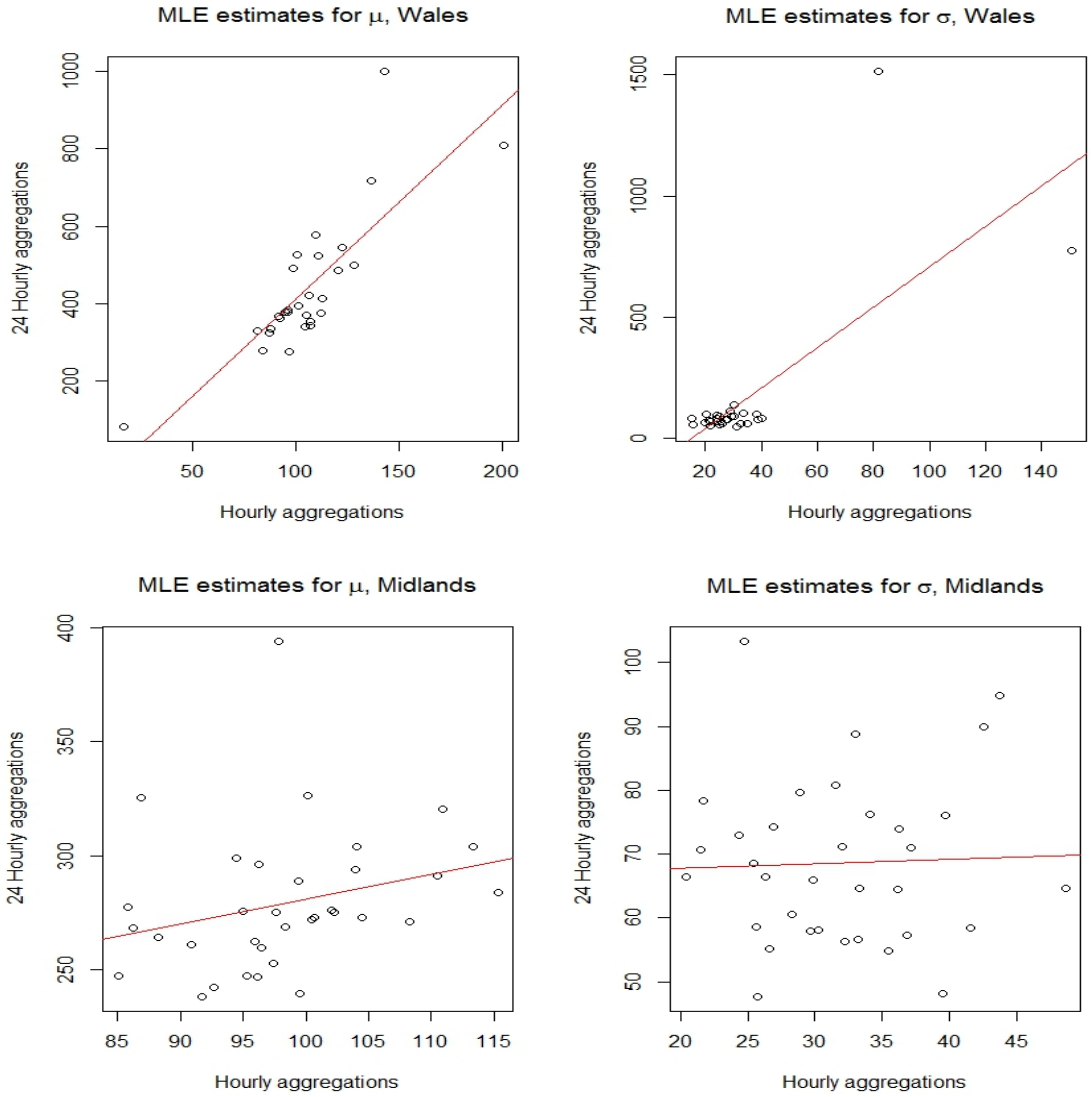

Figure 5.6: MLEs for the GEV parameters using the hourly aggregations plotted against those using the 24 hourly aggregations for Wales and the Midlands.

In contrast to the plots above the MLE estimates for μ and σ in the south East region have high positive correlation, seen in Figure 5.7. There is some clustering in the plot for $\hat{\mu}$ towards the top right of the regression line and clustering in the plot for $\hat{\sigma}$ towards the bottom left of the regression line. It would therefore be feasible to estimate μ and σ using the hourly data data from μ and σ using the 24 hourly aggregations.

The plots for the Anglian region are a little more sparse as this is the region with the

least number of sites used. However, we can still pick out a trend with the plot for $\hat{\mu}$ having fairly high positive correlation. Although this is not so obvious in the plot for $\hat{\sigma}$ we can still see slight positive correlation between the estimates using the hourly data and the 24 hourly aggregations.

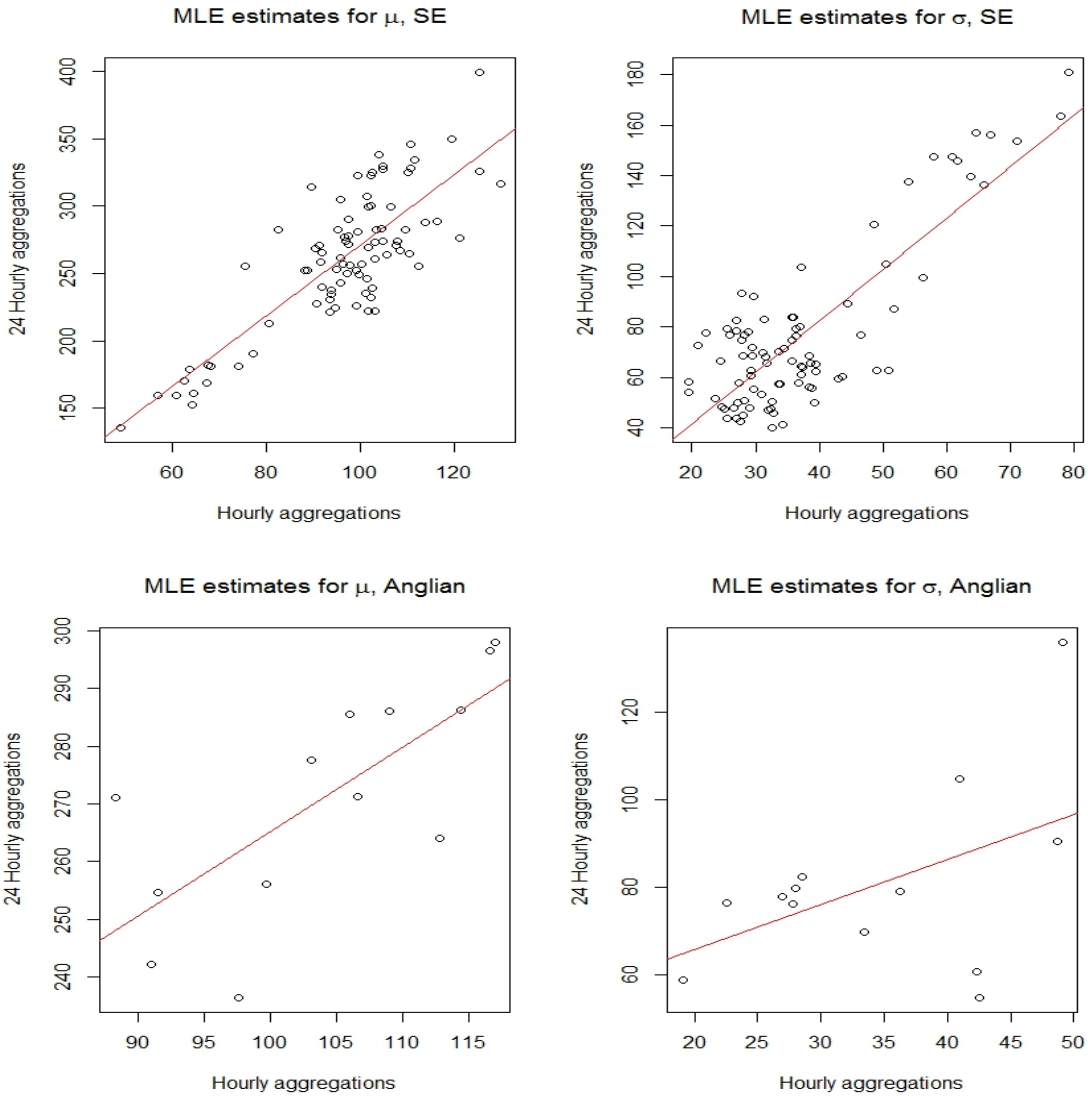

Figure 5.7: MLEs for the GEV parameters using the hourly aggregations plotted against those using the 24 hourly aggregations for the South East and Anglian regions.

Overall, the σ estimates for each region, except the South East, have little or no positive correlation illustrated by a straight line or a line with low gradient. Conversely, the parameter estimates for μ all have high positive correlation except the Midlands where there is only slight positive correlation. This correlation is illustrated by a steep regresion line running throught the majority of the points or the main clusters of points.

Suppose we only had the estimates for each site for μ and σ using the 24 hourly aggregated data. The question we are asking is would we be able to calculate the parameter estimates that would have been found using the hourly data. To do this we would use linear regression and so the red lines that have been drawn on each plot.

Taking the plots for $\hat{\mu}$, the high positive correlation shows a trend in the data and so will make it easier to make inferences about the parameter estimates for the 24 hourly data. However the little or no correlation for $\hat{\sigma}$ would make it difficult to calculate what we think might have happened if we had the hourly data available.

Chapter 6

Return Level Analysis

After fitting each site to the GEV distirbution we can then calculate the return level for each site by the method explained in Chapter 4. Here the 50-year return levels have been calculated as this is a realistic amount that statisticians are interested in. In addition to these, the 1000-year return levels have been calculated. This means we are calculating the chance that a certain amount of rainfall, chosen by the statistician, will happen once every 1000 years.

6.1 The Return Levels

After plotting both the 50-year and 1000-year return levels using the hourly data against the 50-year and 1000-year return levels using the 24 hourly data, for all six regions, it became apparent that for most of them there was little or no pattern. There was an increase in actual values but the stucture of points stayed the same.

Instead of looking at all the plots, explained above, Figure 6.1 shows the two most extreme cases found, the Midlands and the South East region. Firstly looking at the Midlands, we can see a central clustering of points with a regression line fitted to the data cutting nearly horizontally through these points. This suggests that we have little correlation between the return levels calculated with the hourly data and the return levels calculated with the 24 hourly data. This same structure is seen in the 1000-year return levels with a similar central cluster, the only difference being the values at which this cluster is plotted.

The South East region, on the other hand, still has a main cluster of points yet some sites are straying away from this causing the regression line to have a greater gradient. It suggests that a high return level value from the hourly data will result in a higher return level value from the 24 hourly data illustrating a relationship between the two sets of return levels. Analogous to the Midlands region, when we increase the return levels from 50 to 1000 we see a similar structure just with increased values.

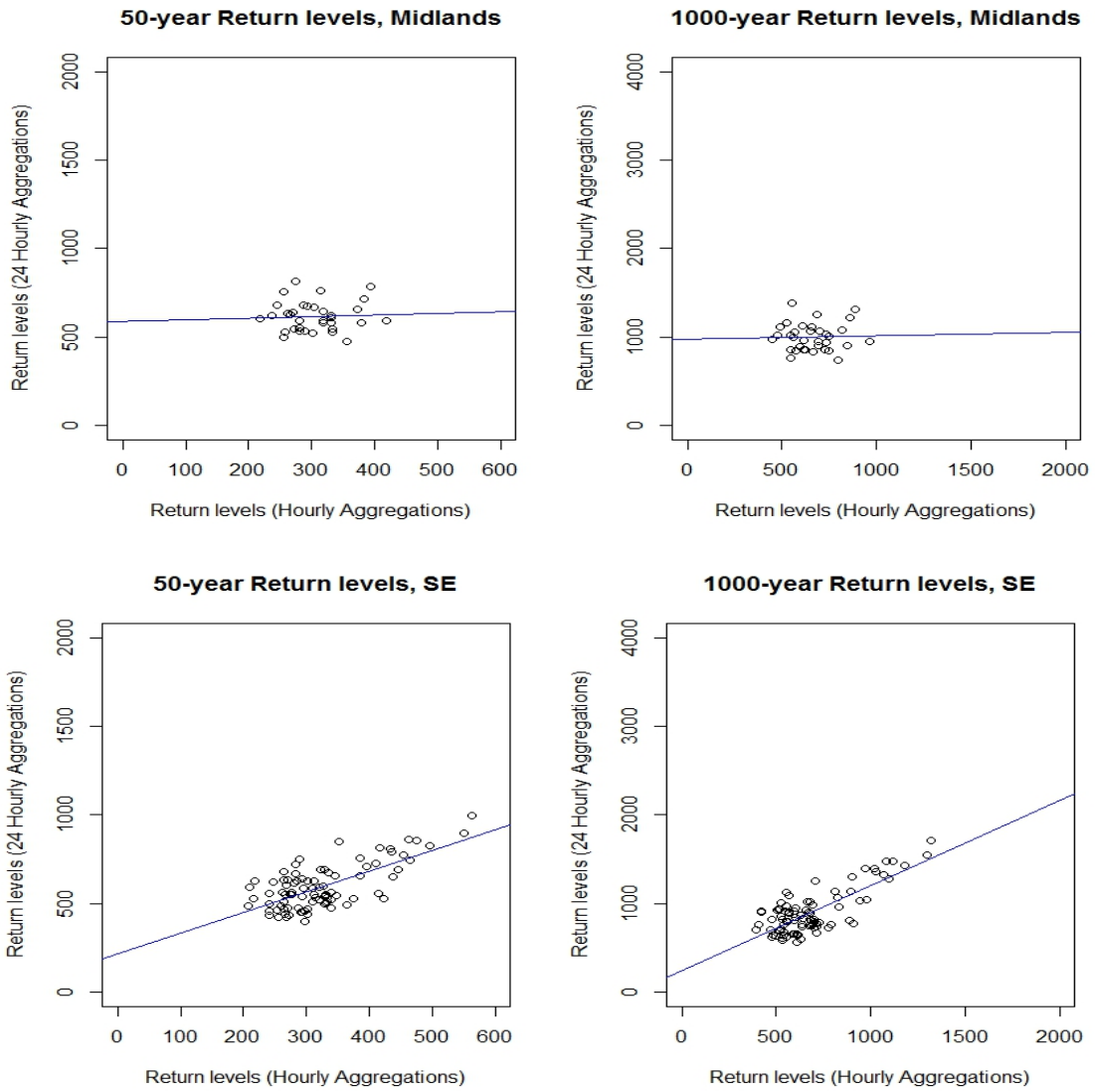

Figure 6.1: 50 and 1000 year Return levels found using hourly data against those found using 24 hourly data for the Midlands and South East regions.

From the 12 return level plots, the 50-year and 1000-year plots for each region, we can conclude that there is no obvious difference in shape between the 50-year and 1000-year plots within a given region; the differences are mainly between the regions themselves. The Midlands are at one end of the scale with zero correlation between the return levels calculated using the hourly data and the return levels calculated using the 24 aggregated data whereas the South East is at the other end of the scale with the highest correlations. The other four regions, are all similar with slight positive correlation with a large amount of clustering within that, hence not all the plots have been shown here.

Using linear regression is only useful when a significant relationship is found between

the variables. Here, when we have zero correlation i.e a horizontal regression line we cannot say anything about what the return levels calculated using hourly data would be if we only had 24 hourly data thus we cannot make inferences about the Midlands.

Looking at the return levels for each region on the same graph we can then see how each region relates to each other. Figure 6.2 below shows both the 50 and 1000-year return levels with each region represented as a different colour. The main difference that stands out is that the 1000-year return levels have a greater spread ranging from approximately 300 to 1200 whereas the 50-year return levels only range from approximately 200 to 600. This is seen with both the hourly and daily data.

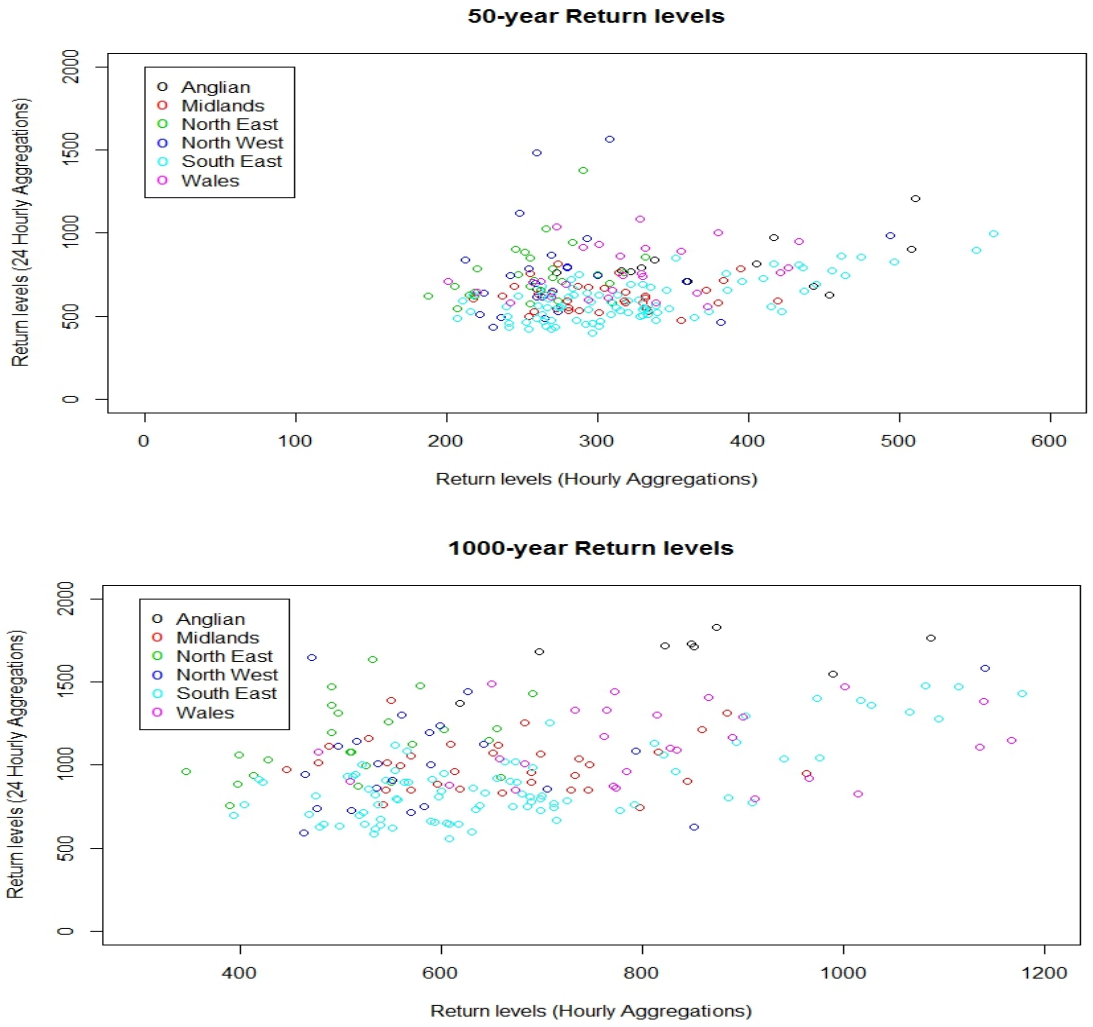

Figure 6.2: 50 and 1000 year Return levels found using hourly data against those found using 24 hourly data for all regions.

In further work we may choose to draw a regression line on both of these plots in Figure

6.2 to show the trend of the points plotted. However, it does not seem sensible at this stage in the analysis due to the large amount of scatter seen on both of the plots, within each region and overall.

6.2 The Standard Errors

Similarly to the MLEs, the return levels are more informative when a standard error is associated with it. However, as $\hat{z}_r$ is now a function of μ, σ and ξ, seen in equation 4.2, we need to use the Delta Method.

This method finds

$$Var(\hat{z}_r) \approx \bigtriangledown {z_r}^T V \bigtriangledown z_r, \tag{6.1}$$

where V is the variance-covariance matrix of the parameter estimates and

$$\bigtriangledown {z_r}^T = \left[\frac{\partial z_r}{\partial \mu}, \frac{\partial z_r}{\partial \mu}, \frac{\partial z_r}{\partial \xi}\right]. \tag{6.2}$$

Taking each element in turn we find that,

$$\begin{aligned}
\frac{\partial z_r}{\partial \mu} &= 1, \\
\frac{\partial z_r}{\partial \sigma} &= \frac{1}{\xi}\left[{y_r}^{-\xi} - 1\right], \\
\frac{\partial z_r}{\partial \xi} &= \frac{\sigma}{\xi}\left(- {y_r}^{-\xi}\log(y_r)\right) + {y_r}^{-\xi}\left(- \frac{\sigma}{\xi}\right) + \frac{\sigma}{\xi} \\
&= \frac{\sigma}{\xi}\left(1 - {y_r}^{-\xi}\right) - \frac{\sigma}{\xi}\left({y_r}^{-\xi}\log(y_r)\right),
\end{aligned}$$

where $y_r = -\log(1 - \frac{1}{r})$, evaluated at $(\hat{\mu}, \hat{\sigma}, \hat{\xi})$. Equation 6.2 then becomes

$$\bigtriangledown {z_r}^T = \left[1, \frac{1}{\xi}\left({y_r}^{-\xi} - 1\right), \frac{\sigma}{\xi}\left(1 - {y_r}^{-\xi}\right) - \frac{\sigma}{\xi}\left({y_r}^{-\xi}\log(y_r)\right)\right] \tag{6.3}$$

Equation 6.1 is calculated using R to give the standard errors for both the 50 and 1000-year return levels for each site. As we have a total of 215 sites, an example of each region has been given below in Table 6.1 for the 50-year return levels and in Table 6.2 for the 1000-year return levels. These would be used in further work to find confidence intervals

Site & Region	RL (Hourly)	s.e.(RL)	RL (Daily)	s.e.(RL)
Ely Common, Anglian	259.7071	127.1495	663.0843	337.2785
Finham, Midlands	319.0088	44.1135	643.7306	69.5375
Harpington, North East	288.7979	35.0306	944.7985	66.4705
Common Bank, North West	254.4373	98.7796	785.6494	161.5276
Princes Marsh, South East	312.9203	26.8405	548.1819	55.0757
Tafalog, Wales	420.8420	67.0502	764.0297	57.1324

Table 6.1: The 50-year Return Levels using the hourly data and the 24 hourly aggregations along with the standard errors for one site in each region.

In the same way that the 95% confidence intervals were calculated in Section 5.1, they can be found here for both the 50-year and 1000-year return levels. However, here we can see that most of the standard errors are too large and so will make the value, in effect, useless.

An example of this is Ely Common in Table 6.2 where the confidence interval will be $(-247.3889, 1484.5063)$. We can see that the range of values is just too big and thus is essentially useless. One explanation for this may be that the number of years used for this region is too small and so we have a higher level of uncertainty reflected in the standard errors.

On the other hand, Harpington, in Table 6.1 has a smaller s.e. for the 50-year return level using the hourly data and so the 95% confidence interval is (220.1379, 357.4579). This seems a more plausible CI to work with which may be caused by a this site having a relatively large amount of data.

Site & Region	RL (Hourly)	s.e.(RL)	RL (Daily)	s.e.(RL)
Ely Common, Anglian	618.5587	441.8100	1369.7660	944.6575
Finham, Midlands	699.7478	139.6341	1067.1867	162.8554
Harpington, North East	679.7511	116.3131	1478.5416	181.2851
Common Bank, North West	515.6715	276.9549	1144.7740	312.3342
Princes Marsh, South East	644.1190	82.4928	829.5273	110.2217
Tafalog, Wales	1135.1511	305.3405	1105.9008	147.2666

Table 6.2: The 1000-year Return Levels using the hourly data and the 24 hourly aggregations along with the standard errors for one site in each region.

6.3 Depth-duration-frequency curve

As mentioned in Chapter 1, we can plot several return levels for each time aggregations, namely a depth-duration-frequency curve. An example of these can be seen below in Figure 6.3 where we have taken Jesmond from the North East.

The structure of the plot is as expected with the return levels for the daily data consistently higher than the return levels for the six hourly and hourly data. We can see a slow increase for the daily data which if we increased the return period we would expect it to continue. The hourly and six hourly return levels on the other hand are slowly increasing but it looks like they will both eventually dip as the gradient if we were to increase the return period anymore.

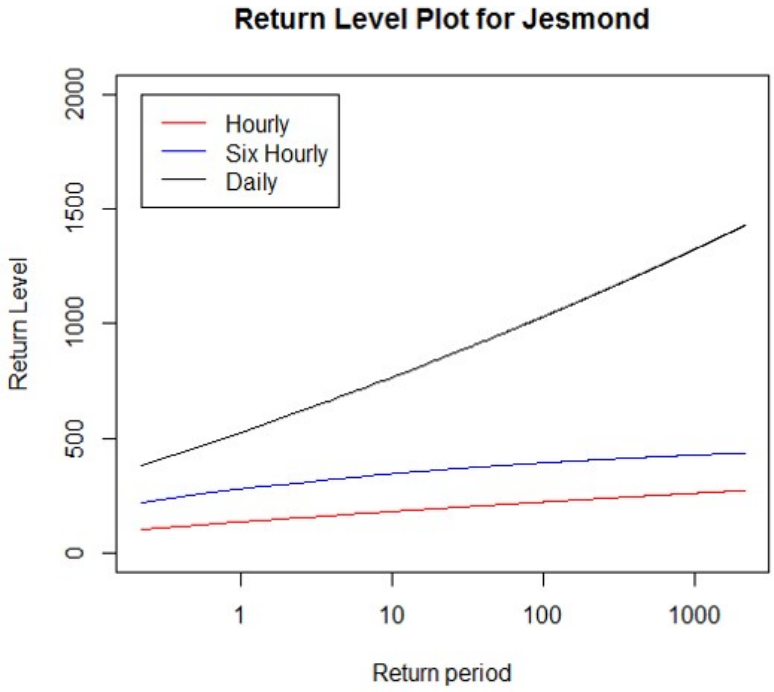

Figure 6.3: A depth-duration-frequenct curve for Jesmond, NE.

Chapter 7

Generalised Likelihood Ratio Test

The calculations in the above Chapters, including the return levels and the regression lines, have all been found by assuming a common ξ throughout each region. This is thought to be appropriate due to the sites in each region having a similar geographical structure.

However, we will now investigate if this was a correct assumption my comparing the log-likelihoods obtained by each method. Let the model with varying ξ be Model A with p covariates, and the model with common ξ be Model B with q covariates and so Model B is nested within Model A i.e. $q < p$. We are then testing these against each other i.e.

$$H_0 : \text{Model } B$$
$$H_1 : \text{Model } A$$

A standard generalised Likelihood Ratio test is then performed, that is we take twice the difference between the maximised log-likelihood under each model and test this value against the Chi-squared distribution where the degrees of freedom is equal to the difference between the number of parameters in each model. Let L_A be the likelihood from Model A and L_B be the likelihood from Model B.

Table's 7.1 and 7.2 below show the test statistics we have calculated for each region for both Model 1 and Model 2 when using the hourly data and the daily data. Here, all the values are negative as the R programme used, finds the negative log-likelihood. This not a probem as both L_A and L_B are the negative log-likelihoods.

It can also be noticed that for Model A the South East has not converged when using the daily data and using the hourly data both the South East and Wales have not converged. In further work, this would be investigated but in this report we will treat it as miss-calculations in R due to values oscillating rather than converging.

Region	L_A	L_B	$2(L_A\text{-}L_B)$
Anglian	-1124.249	-1130.899	13.3
Midlands	-3987.016	-3998.162	22.929
North East	-2706.075	-2731.640	49.131
North West	-2101.879	2116.706	29.654
South East	Not converged	-8756.441	-
Wales	Not converged	-2751.22	-

Table 7.1: The maximised log-likelihood for each region for both Models, using the hourly data.

If we were to just look at Table 7.1 we can compare each test statsitic against the corresponding value from the Chi-squared distribution, that is,

$$2(L_A - L_B) \sim \chi^2_{p-q}.$$

This gives that the only significant region is the North East, thus we reject H_0 in favour of H_1 and so for this region Model A is more appropriate suggesting we should let ξ vary over the region. For the other three regions we have values for, our assumption to keep a common ξ within each region holds.

However, when we look at Table 7.2 our findings above are contrasted and for the North East we would accept H_0 suggesting that a common ξ over the region is appropriate. The North West and Wales on the other hand are significant and so for these regions we would reject H_0 assuming that letting ξ vary over the region is suitable for this data.

Region	L_A	L_B	$2(L_A\text{-}L_B)$
Anglian	-1129.109	-1141.459	24.698
Midlands	-4451.496	-4473.012	43.032
North East	-3248.206	-3259.464	22.516
North West	-2499.100	-2521.854	45.508
South East	Not converged	-9675.941	-
Wales	-3104.534	-3187.243	165.418

Table 7.2: The minima for each region for both Models, using the daily data.

Chapter 8

Conclusion

The aim of this paper was to build a model for the annual maximum rainfall at 215 locations across England and Wales split into six regions. We were given the hourly data but it was then aggregated by 3, 6, 12 and 24 hours to given four new data sets for each site. The problem was approached by fitting the data regionally to a GEV distribution and thus obtaining location, scale and shape parameter estimates, μ, σ and ξ respectively. The question then asked was if we only had the daily data could we predict what would happen for the hourly data? This question was considered later in the paper.

Due to within region, geographical similarities we started with fixing ξ for each region and then fitting the model. However, we then checked if this asumption was correct by letting ξ vary from site to site and then comparing these models using a Likelihood Ratio Test.

The annual maxima were extracted using the statistical programme R which were then fed through a function that would find the log likehood and parameter estimates for each region. Also in R the 50-year and 1000-year return levels were calculated, along with their standard errors, and plotted to compare each region.

Over the six regions the MLEs for μ, σ and ξ were found and plotted to look for any patterns. For each region we found that as we increase the time aggregations both $\hat{\mu}$ and $\hat{\sigma}$ increase whereas $\hat{\xi}$ appears to decrease. Looking further into the values for $\hat{\xi}$ in Table 5.1 we saw that indeed they do decrease as the time aggregations are increased.

Further analysis of the parameter estimates led to linear regression performed for each region. We saw that for most of the regions there was slight positive correlation when the MLEs based on the 24 hourly aggregations were plotted against the hourly data. However, for Wales, the South East and Anglian regions we saw high positive correlation. The regression lines then enable us to attempt to answer the question that was asked at the beginning. We would indeed be able to say something about the MLEs found by using the hourly data even if we only had the daily data.

The 50-year and 1000-year return levels were then compared for each region with the two extremes being the Midlands and the South East. The Midlands has a central cluster of points with little correlation whereas the South East has positive correlation i.e. as the return levels for the hourly aggregations increase so do the return levels for the 24 hourly aggregations. The return levels for each region were then plotted on a single graph to show how the spread of points increases from 50-year return levels to the 1000-year return levels. Similar to the MLEs, the standard errors were calculated for each site using the Delta Method with an example from each region given in Table 6.1 for the 50-year return levels and in Table 6.2 for the 1000-year return levels.

As mentioned previously, we assumed a common ξ throughout each region when proceeding with the above methods, however how do we know if this assumption was correct? To test this we fitted the same model to the data but this time letting ξ vary from site to site giving the likelihoods in Table 7.1 those found using the hourly data and the likelihoods in Table 7.2 found using the daily data. A likelihood ratio test was then performed, on all the regions possible, concluding that when we use the hourly data it is correct to assume a common ξ for Anglian, the Midlands and the North West. However, our assumption was challenged for the North East as we found that letting ξ vary throughout each region was more appropriate here.

When using the daily data, we found that our assumption of a common ξ was correct for Anglian, the Midlands and the North East but for the North West and Wales a seperate ξ for each site is more appropriate.

One of the problems encountered when carrying out the analysis was the varying number of observations for each region due to the quality control constraint. This may mean that some regions have a higher uncertainty attributed to them which would be addressed in further work.

Possible extensions to this analysis would be to investigate the assumption of a common ξ further with a deeper explanantion of why some regions were not converging. In addition depth-duration-frequency curves could be explored for a greater number of regions with some predictions of what might happen if the hourly data was unavailable.

Bibliography

[1] Casella, G. and Berger, R. L. (2002), Statistical Inference

[2] Coles, S. (2001), An Introduction to Statistical Modeling of Extreme Values.

[3] Davison, A. C. (2003), Statistical Models

[4] Fawcett, L. (2012), MAS8391 Modelling Environmental Extremes.

[5] Fisher, R.A.; Tippett, L.H.C. (1928), "Limiting forms of the frequency distribution of the largest and smallest member of a sample"

[6] Gnedenko, B.V. (1943), "Sur la distribution limite du terme maximum d'une serie aleatoire"

[7] Gumbel, E.J. (1958), Statistics of Extremes

[8] Katz, R. (), Extreme Value Analysis for Climate Series
https://www.isse.ucar.edu/staff/katz/docs/pdf/banffrwk.pdf

[9] R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Appendix

Listing 8.1: Selecting the maxima for each region

```
#the following function takes each region, then within that site, and selects
#the annual maximum for the hourly, 3 horly, 6 hourly, 12 hourly and daily
#aggregations and stores them in a list with a name for each region

selectmax=function(region,list){

  sites=list()
  for (k in 1:length(list))
  {
    name=list[k]
    #target=paste("E:\\sites\\",region,"\\",name,".txt",sep="")
    target=paste("F:\\sites\\",region,"\\",name,".txt",sep="")
    print(target)
    data=read.table(target)
    newdata=data.frame(data)
    colnames(newdata)=c("A","B","Year","Month","Day","Hour","Rainfall")
    t=vector()
    t=c(1,3,6,12,24)
    n=63*8760+15*24
    rep=list()
    sum=list()
    yrs=list()
    max=list()
    ind=list()
    missvals=list()
    allvals=list()
    maxima=list()


    for(i in 1:5){
      rep[[i]]=rep(1:(n/t[i]),each=t[i])
      sum[[i]]=tapply(newdata$Rainfall,rep[[i]],sum)
      yrs[[i]]=newdata$Year[(1:nrow(newdata) %% t[i])==0]
      max[[i]]=tapply(sum[[i]],yrs[[i]],max)
      ind[[i]]=sum[[i]]<(-1)
      missvals[[i]]=tapply(ind[[i]],yrs[[i]],sum)
      allvals[[i]]=tapply(ind[[i]],yrs[[i]],length)
      maxima[[i]]=tapply(sum[[i]],yrs[[i]],max)[missvals[[i]]/allvals[[i]]<0.15]
      #       sites[[k]][[i]]=maxima[[i]]
    }
    sites[[k]]=maxima
  }
  return(sites)
}
```

Listing 8.2: The log-likelihood and MLEs keeping a common ξ

```
#The following take the list of annual maximums for each region and runs them
#through an iterative function to calculate the -loglikelihood. The parameter
#estimates are also found with this function keeping a common xi throughout
#each region.

k=5
theta=vector()
s=27
for(i in 1:s){
  theta[i]=mean(dataset1[[i]][[k]])
  theta[s+i]=sd(dataset1[[i]][[k]])
}
theta[2*s+1]=0.1


gev.loglike=function(theta){

  mu=vector()
  sigma=vector()
  for(i in 1:s){
    mu[i]=theta[i]
    sigma[i]=theta[s+i]
    xi=theta[2*s+1]
  }
  loglike=vector()
  a=vector()
  for(i in 1:s){
    a[i]=min((1+(xi*(dataset1[[i]][[k]]-mu[i])/sigma[i])))
  }
  if(min(a)<0.00001)return(1000000)
  if(min(sigma)<0.00001)return(1000000)
    for(j in 1:s){
      loglike[j]=-length(dataset1[[j]][[k]])*log(sigma[j])-(1/xi+1)
          *sum(log(1+(xi*(dataset1[[j]][[k]]-mu[j])/sigma[j])))-
          sum((1+(xi*(dataset1[[j]][[k]]-mu[j])/sigma[j]))^(-1/xi))
    }

    logliketot=sum(loglike)
  return(-logliketot)
}
```

Listing 8.3: The log-likelihood and MLEs with varying ξ

```
#The following take the list of annual maximums for each region and runs them
#through an iterative function to calculate the -loglikelihood. The parameter
#estimates are also found with this function varying xi throughout each region.

k=1
theta=vector()
s=29
for(i in 1:s){
  theta[i]=mean(dataset1[[i]][[k]])
  theta[s+i]=sd(dataset1[[i]][[k]])
  theta[2*s+i]=0.1
}

gev.loglike.xi=function(theta){

  mu=vector()
  sigma=vector()
  xi=vector()
  for(i in 1:s){
    mu[i]=theta[i]
    sigma[i]=theta[s+i]
    xi[i]=theta[2*s+i]
  }
  loglike=vector()
  a=vector()
  for(i in 1:s){
    a[i]=min((1+(xi[i]*(dataset1[[i]][[k]]-mu[i])/sigma[i])))
  }
  if(min(a)<0.00001)return(1000000)
  if(min(sigma)<0.00001)return(1000000)
    for(j in 1:s){
      loglike[j]=-length(dataset1[[j]][[k]])*log(sigma[j])-(1/xi[j]+1)*
          sum(log(1+(xi[j]*(dataset1[[j]][[k]]-mu[j])/sigma[j])))
          -sum((1+(xi[j]*(dataset1[[j]][[k]]-mu[j])/sigma[j]))^(-1/xi[j]))
    }

    logliketot=sum(loglike)

  return(-logliketot)
}
```

Listing 8.4: Using the delta method to find the standard errors of a return level

```
#the following code is an example for the Anglian region using the delta method to
#find the 50-year return level standard errors using hourly and daily data

y50=-log(1-(1/50))
#######anglian
###hourly se
angdel1.xi=vector()
for(i in 1:13){
  angdel1.xi[i]=((angmle1[i+13])*((angmle1[27])^(-2))*(1-((y50)^(-angmle1[27]))))
  -((angmle1[i+13])*((angmle1[27])^(-1))*((y50)^(-(angmle1[27])))*log(y50))
}

angdel1.mu=1
angdel1.sigma=-((angmle1[27])^(-1))*(1-(y50^(-angmle1[27])))

angdel1.s=matrix(ncol=26,nrow=13,0)
for(i in 1:13){
  angdel1.s[i,2*i-1]=angdel1.mu
  angdel1.s[i,2*i]=angdel1.sigma
}

angdel.1=cbind(angdel1.s,angdel1.xi)

ang.rl.se.1=diag(sqrt(angdel.1%*%angv1%*%t(angdel.1)))

###24 hourly se
angdel24.xi=vector()
for(i in 1:13){
  angdel24.xi[i]=((angmle24[i+13])*((angmle24[27])^(-2))*(1-((y50)^(-angmle24[27]))))
  -((angmle24[i+13])*((angmle24[27])^(-1))*((y50)^(-(angmle24[27])))*log(y50))
}

angdel24.mu=1
angdel24.sigma=-((angmle24[27])^(-1))*(1-(y50^(-angmle24[27])))

angdel24.s=matrix(ncol=26,nrow=13,0)
for(i in 1:13){
  angdel24.s[i,2*i-1]=angdel24.mu
  angdel24.s[i,2*i]=angdel24.sigma
}

angdel.24=cbind(angdel24.s,angdel24.xi)

ang.rl.se.24=diag(sqrt(angdel.24%*%angv24%*%t(angdel.24)))
```