Statistics for Environmental Evaluation:
Quantifying the environment

Extreme value modelling: Answer Sheet

1 Data set

The data set `rain` is a vector of length 17351 containing daily observations of rainfall accumulations in mm at a location in South West England.

```
data(rain)
plot(rain)
```

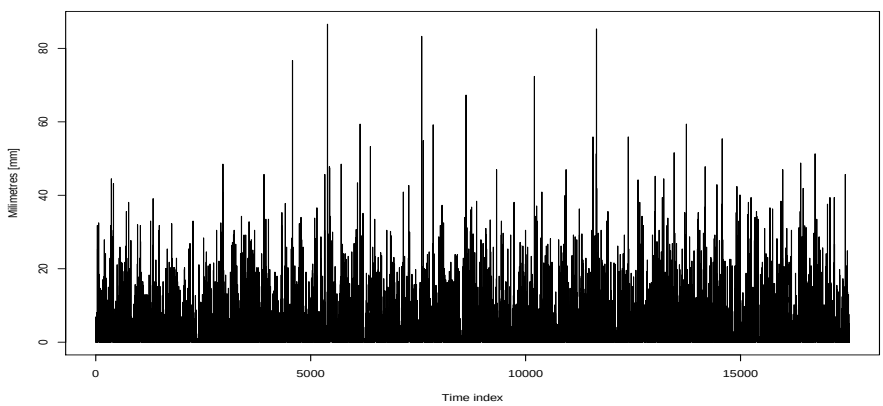

Figure 1: Daily observations of rainfall in millimetres at a location in South West England

There are no obvious trends in the data and a threshold of at least 20mm should be chosen in order to ensure that we are in the tail of the data set.

2 Fitting the GPD to threshold excesses - rain data

Examine the choice of threshold required to fit the GPD to this data set.

2.1 Threshold diagnostics

The mean residual life plot `mrl.plot` and parameter stability plot `gpd.fitrange` are used to determine the level at which to set the threshold u.

```
mrl.plot(rain)
gpd.fitrange(rain,umin=20, umax = 50)
```

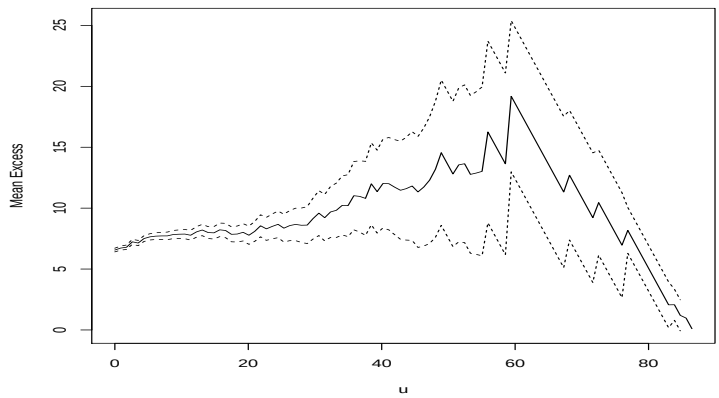

Figure 2: Mean residual life plot

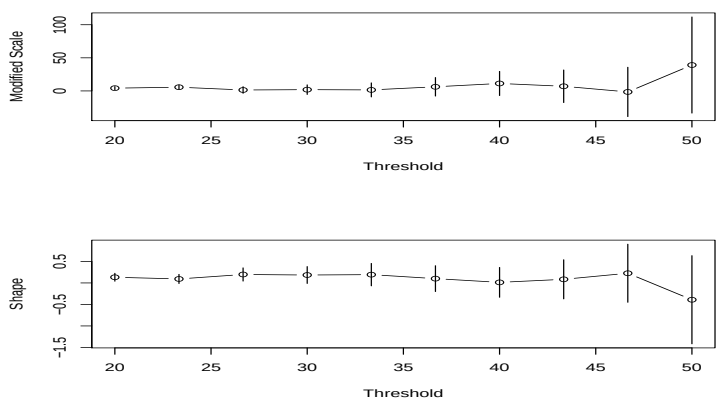

Figure 3: Parameter stability plot

We chose a threshold equal to $20mm$.

2.2 Fitting the GPD to the threshold excesses

2.2.1 Fitting the GPD

```
rain.fit<-gpd.fit(rain, threshold = 20)
1-rain.fit$rate
```

We have chosen a threshold equal to $20mm$, which corresponds to a 97% quantile.

Note: It is worth fitting the generalised Pareto distribution to data with a number of different thresholds to see how this effects the estimate of the parameters as well as the model fit.

2.2.2 Parameter estimates

1. What are your estimates of the GPD parameters, and what are their standard errors and their corresponding 95% confidence intervals?

Scale parameter (σ) estimate and standard error: 6.83 (0.43) respectively.

```
rain.fit$mle[1];rain.fit$se[1]
```

95% confidence interval for the scale parameter: $(5.98, 7.68)$

```
c(rain.fit$mle[1]-1.96*rain.fit$se[1],rain.fit$mle[1]+1.96*rain.fit$se[1])
```

Shape parameter (ξ) estimate and standard error: 0.13 (0.04) respectively

```
rain.fit$mle[2];rain.fit$se[2]
```

95% confidence interval for the scale parameter: $(0.04, 0.23)$

```
c(rain.fit$mle[2]-1.96*rain.fit$se[2],rain.fit$mle[2]+1.96*rain.fit$se[2])
```

2.2.3 Diagnostic plot

The following command can be used to assess the fit of the GPD.

```
gpd.diag(rain.fit)
```

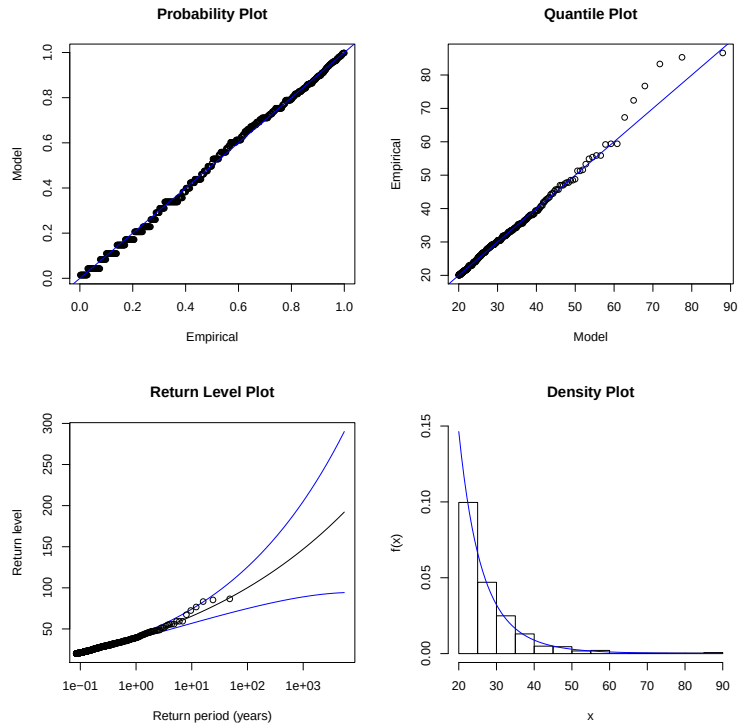

Figure 4: Diagnostic plot of the generalised Pareto distribution for the rainfall data set.

The return values can be read off of the return level plot to give an estimate of 90 with 95% confidence interval (70, 130). Remember these return values are conditional on being above the threshold, therefore when reporting return values we would want to account for the rate at which we see observations above the threshold.

2.3 Fitting the GPD to threshold excesses with covariates

Using the threshold chosen in the previous analysis examine the evidence for a linear trend in the scale parameter.

$$\sigma_u(t) = \exp(\alpha + \beta t)$$

The covariate `t` can be created by using the following command:

```
t<-seq(0,1,length.out=length(rain)).
```

Use the documentation for the `gpd.fit` function to find out how to use a logarithmic link to ensure positivity of the fitted $\sigma_u(t)$.

2.3.1 Does a likelihood ratio test suggest any evidence for a linear trend in the scale parameter?

Model fitting:

```
rain.fit2<-gpd.fit(rain,threshold=20,ydat=matrix(t,ncol=1),sigl=1,siglink=exp)
```

Performing the likelihood ratio test:

```
rain.fit.llh<--1*rain.fit$nllh

rain.fit2.llh<--1*rain.fit2$nllh

dev.value<-2*(rain.fit2.llh-rain.fit.llh)

p.val<-1-pchisq(q=dev.value,df=1)
```

The p-value obtained in the likelihood ratio test to compare models `rain.fit` and `rain.fit2` is 0.226. Therefore, the linear time trend in the scale parameter is not statistically significant.

2.3.2 Does a likelihood ratio test suggest any evidence for a linear trend in the shape parameter?

Model fitting:

```
rain.fit3<-gpd.fit(rain,threshold=20,ydat=matrix(t,ncol=1),shl=1,shlink=identity)
```

Performing the likelihood ratio test:

```
rain.fit.llh<--1*rain.fit$nllh

rain.fit3.llh<--1*rain.fit3$nllh

dev.value2<-2*(rain.fit3.llh-rain.fit.llh)

p.val<-1-pchisq(q=dev.value2,df=1)
```

The p-value obtained in the likelihood ratio test to compare models `rain.fit` and `rain.fit3` is 0.850. Therefore, the linear time trend in the shape parameter is not statistically significant.

2.3.3 Are there any other covariates that you think it would be useful to have information about?

- Information on the North Atlantic Oscillation (NAO).
- Information on when in the year the rain occurred - are we seeing more summer or winter rainfall events?
- There will be many more covariates that could be potentially informative about extreme rainfall.