Statistics for Environmental Evaluation: Quantifying the environment

Extreme value modelling: Worksheet

Your aim for this workshop is to learn how to use RStudio to fit the generalised Pareto distribution to an environmental data set.

Upon completing this workshop you should be able to:

- Visualise a time series of environmental data
- Use diagnostic tools to determine a suitable threshold for the data set
- Fit a generalised Pareto distribution to the data set
- Assess the model fit and behaviour of return values
- Incorporate covariates into the generalised Pareto distribution

1 Recap of the generalised Pareto distribution

The generalised Pareto (GP) distribution is conditional on a random variable X being above a predetermined threshold u.

$$\mathbb{P}\left(X \leq x | X > u\right) \;=\; 1 - \left[1 + \xi\left(\frac{x-u}{\sigma_u}\right)\right]_+^{-1/\xi} \qquad \text{for} \qquad x_1, \ldots, x_n,$$

where σ_u is the scale parameter of the distribution, which is dependent on the threshold u. The shape parameter ξ tells us about the tail of the distribution.

2 R packages

The R package `ismev` is the standard package for fitting an extreme value model to a single time series. All the necessary functions are contained in the library `ismev`.

```
install.packages("ismev")
library(ismev)
```

`??ismev` This command provides you with more information about the R functions that are available in the library.

3 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)
```

From plotting the data set, think about the following questions:

- Are there any trends visible in the data set?
- From a visual inspection at what level should we pick our threshold for fitting the GPD?

4 Fitting the GPD to threshold excesses - rain data

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

4.1 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= YOUR CHOICE, umax = YOUR CHOICE)
```

4.2 Fitting the GPD to the threshold excesses

Once you have determined a threshold we can now fit the generalised Pareto distribution.

```
rain.fit<-gpd.fit(rain, threshold = YOUR THRESHOLD)
```

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

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

For your choice of threshold, consider the following questions:

1. Why did you choose this particular threshold?
2. What are your estimates of the GPD parameters, and what are their standard errors?
3. What do the estimate of the shape parameter and its 95% confidence interval tell you about the tail of the fitted distribution?
4. What do your diagnostic plots tell you about the quality of your fit?
5. Use the diagnostic plot to report a 95% confidence interval for the 100 year return value and compare this with levels observed in your data.

4.3 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 $\sigma_u(t)$.

- Does a likelihood ratio test suggest any evidence for a linear trend in the scale parameter?
- Does a likelihood ratio test suggest any evidence for a linear trend in the shape parameter?
- Are there any other covariates that you think it would be useful to have information about?