Solution 1 To run the code we load the packages:

```
load(evd, mev, ismev, scales, lubridate, gridExtra, ggplot2, dplyr, tidyr, ggdist,
 ggpubr, xts)
```

The following code gives Figure 1.

```
load("eskrain.RData")

time.seq <- seq(from=min(date(eskrain)), to=max(date(eskrain)), length=149016)
precip_numeric <- as.numeric(eskrain)
esk.rain <- data.frame(date=as.Date(time.seq), precip=precip_numeric)

u <- 5
plot_esk <- plot(esk.rain, type="h", ylab="Hourly rainfall (mm)", xlab="Time")
points(esk.rain[esk.rain$precip > u,], col="red", cex=.25, pch=20)
abline(u, 0, col="red")
```

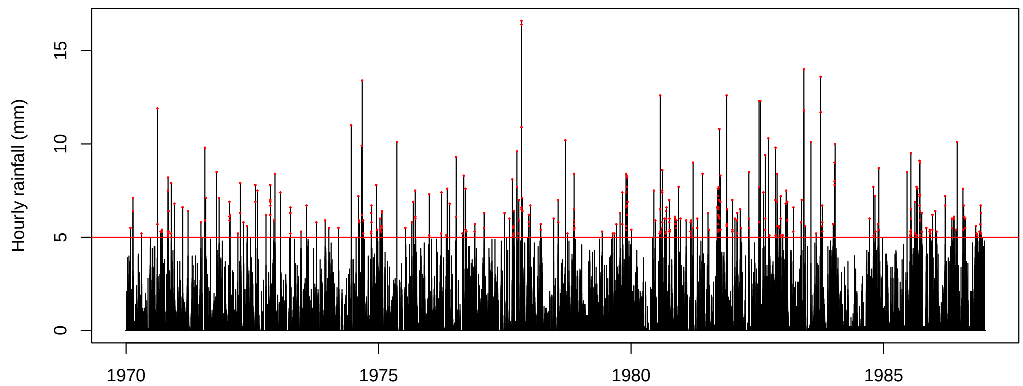

Figure 1: Precipitation from 1970–1986 in Eskdalemuir.

(a) We run the following to find the maxima for the different periods:

```
### start by taking daily maxima
daily.max <- apply(matrix(esk.rain$precip, ncol=24, byrow=T), max)

### then use daily maxima to compute weekly maxima
weekly.max <- apply(matrix(daily.max, ncol=7, byrow=T), max)

### finally, take monthly maxima. We use here months with fixed lengths of 30 days
monthly.max <- apply(matrix(daily.max, ncol=30, byrow=T), max)
```

There are many dry days with zero precipitation, giving a positive probability mass at zero, and this will impact the fit of the GEV to daily data. For instance, analysis of the daily maxima results in a warning on the convergence of the optimisation routine, and gives the following fit

```
(fit.daily <-  fgev(daily.max))

Call: fgev(x = daily.max)
Deviance: 10475.06

Estimates
    loc    scale    shape
0.02312  0.05478  2.19920

Standard Errors
      loc      scale      shape
2.025e-06  1.082e-03  7.285e-02

Optimization Information
  Convergence: iteration limit reached
  Function Evaluations: 1184
  Gradient Evaluations: 100
```

The diagnostic plots in Figure 2 show that the fitted model does not describe the data well.

```
par(mfrow=c(1, 4))
plot(fit.daily)
```

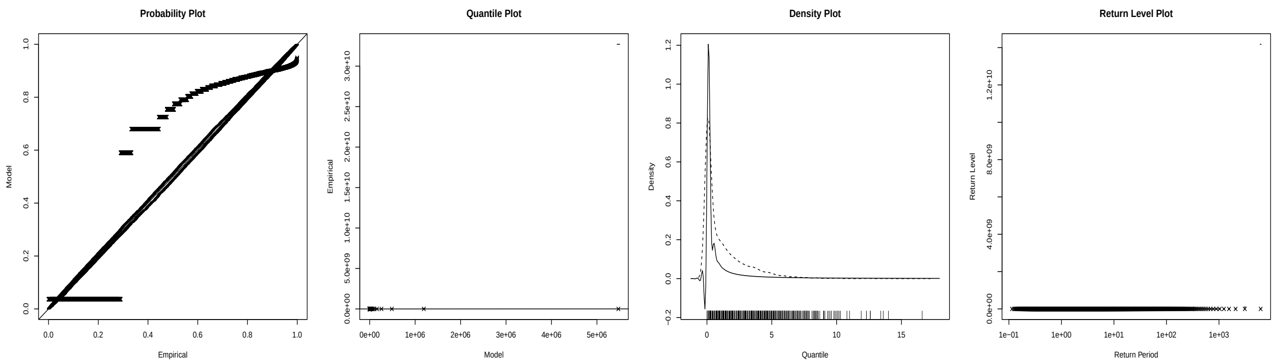

Figure 2: Diagnostic plots of the GEV fit for daily maxima.

Fitting the GEV model to the weekly maxima gives a much better description of the data. The MLEs are $\widehat{\nu} = 2.25$, $\widehat{\sigma} = 1.83$ and $\widehat{\xi} = 0.05$. This is confirmed by the plots in Figure 3. Nonetheless, the probability plots reveal that the fit to the bulk of the distribution is not ideal.

```
(fit.weekly <-  fgev(weekly.max))

Call: fgev(x = weekly.max)
Deviance: 3934.709

Estimates
    loc    scale    shape
2.25220  1.83398  0.05482

Standard Errors
    loc    scale    shape
0.07141  0.05403  0.02997
```

	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\xi}$
Weekly	2.25 (0.07)	1.83 (0.05)	0.05 (0.03)
Monthly	4.78 (0.16)	2.05 (0.11)	-0.02 (0.05)

Table 1: Parameter estimates for the fitted GEV models.

```
par(mfrow=c(1, 4))
plot(fit.weekly)
```

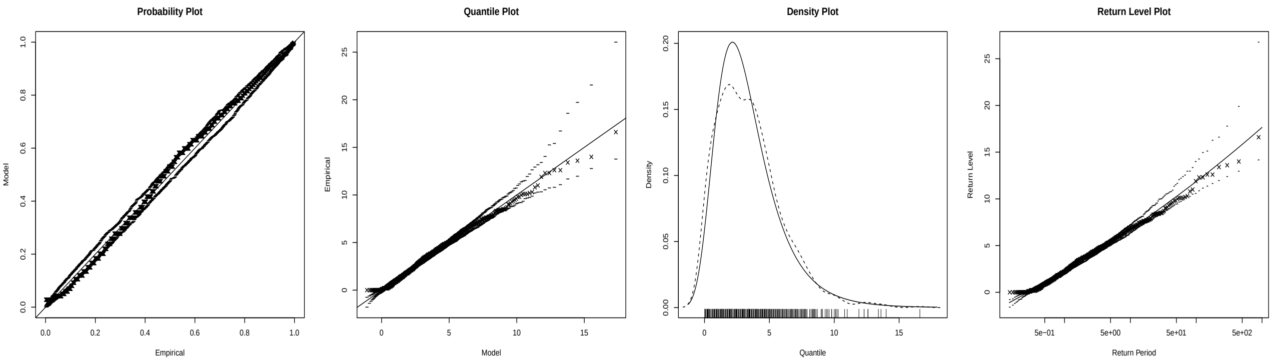

Figure 3: Diagnostic plots of the GEV fit for weekly maxima.

Finally, we analyse the monthly maxima. The resulting MLEs are $\widehat{\nu} = 4.78, \widehat{\sigma} = 2.05$ and $\widehat{\xi} = -0.02$. The diagnostic plots in Figure 4 show a good fit.

```
(fit.monthly <-  fgev(monthly.max))

Call: fgev(x = monthly.max)
Deviance: 944.4168

Estimates
     loc     scale     shape
 4.78404   2.04649  -0.02086

Standard Errors
    loc    scale    shape
0.15925  0.11445  0.04841

par(mfrow=c(1, 4))
plot(fit.monthly)
```

(b) We only look at the models for the weekly and monthly maxima. Table 1 gives the MLEs and their standard errors (in brackets). Suppose that we want to compute 0.95% confidence intervals (CIs). We use standard likelihood theory (e.g., slide 24) to compute the CIs with confidence level $1 - 2\alpha$. Here $\alpha = 0.025$, so $z_{1-\alpha} = 1.96$ and then obtain the CIs in Table 2, using for instance

```
fit.monthly$estimate+qnorm(0.975)*c(-1,1)*fit.monthly$std.err
```

These CIs are based on the normal approximation to the distribution of the estimates, but we should check whether this is reasonable using the profile log likelihood plots to see whether there are asymmetries:

```
par(mfrow=c(1,3))
plot(profile(fit.monthly))
```

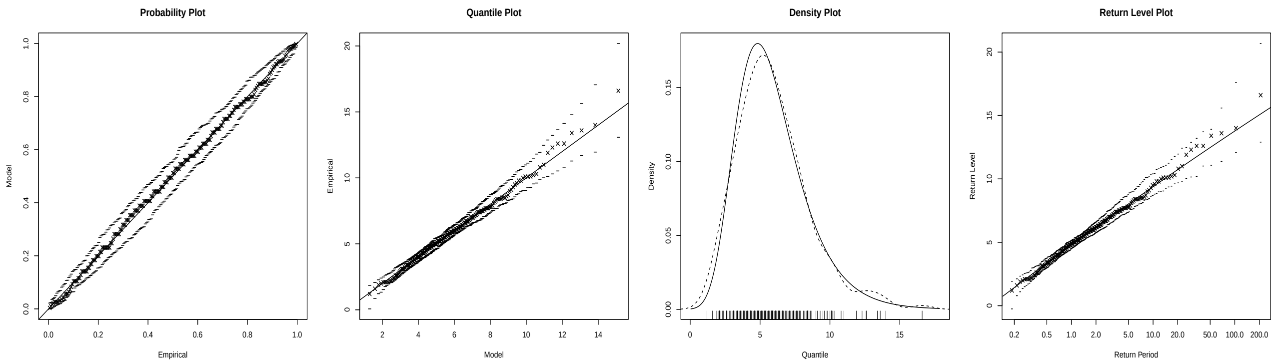

Figure 4: Diagnostic plots of the GEV fit for monthly maxima.

	Weekly	Monthly
$\widehat{\mu}$	(2.11, 2.39)	(4.47, 5.10))
$\widehat{\sigma}$	(1.73, 1.94)	(1.82, 2.27)
$\widehat{\xi}$	(-0.004 , 0.114)	(-0.12, 0.07)

Table 2: 95% confidence intervals for the GEV parameters.

Figure 5 shows fairly symmetric profiles, so the confidence intervals above should be reasonable, except maybe for the shape parameter.

(c) The confidence interval for ξ in Table 2 does not exclude the possibility that $\xi = 0$. For a test based on the likelihood ratio statistic between a GEV model with three parameters and the Gumbel model, i.e., the GEV model with only location and scale parameters, we can use the difference of deviances for the two models:

```
fit.monthly0 <- fgev(monthly.max, shape=0)
ratio <- fit.monthly0$dev - fit.monthly$dev)
qchisq(0.95,1)
3.841459
1-pchisq(ratio,df=1)
0.672707
```

As the p-value equals 0.67 and `ratio` does not exceed the theoretical 95% quantile of a χ^2_1 random variable, we cannot reject the Gumbel model under which $\xi = 0$.

Solution 2 We now study the excess precipitation above a threshold u for the Eskdalemuir rain data.

(a) We start by looking at the mean residual life plot

```
mrlplot(esk.rain$precip)
```

Figure 6 shows that the mean excesses exhibit a rather stable 'linear' behaviour for $2 \leq u \leq 5$, see also slide 92. There is a slight upward trend in the figure, suggesting that $\xi > 0$. These impressions are confirmed by the parameter stability plots in Figure 7, which show stable behaviour of the estimates for such values of u (though they are less suggestive that $\xi > 0$):

```
tcplot(esk.rain$precip, tlim=c(0.1,7), model="gpd", nt = 20)
```

(b) We run the code

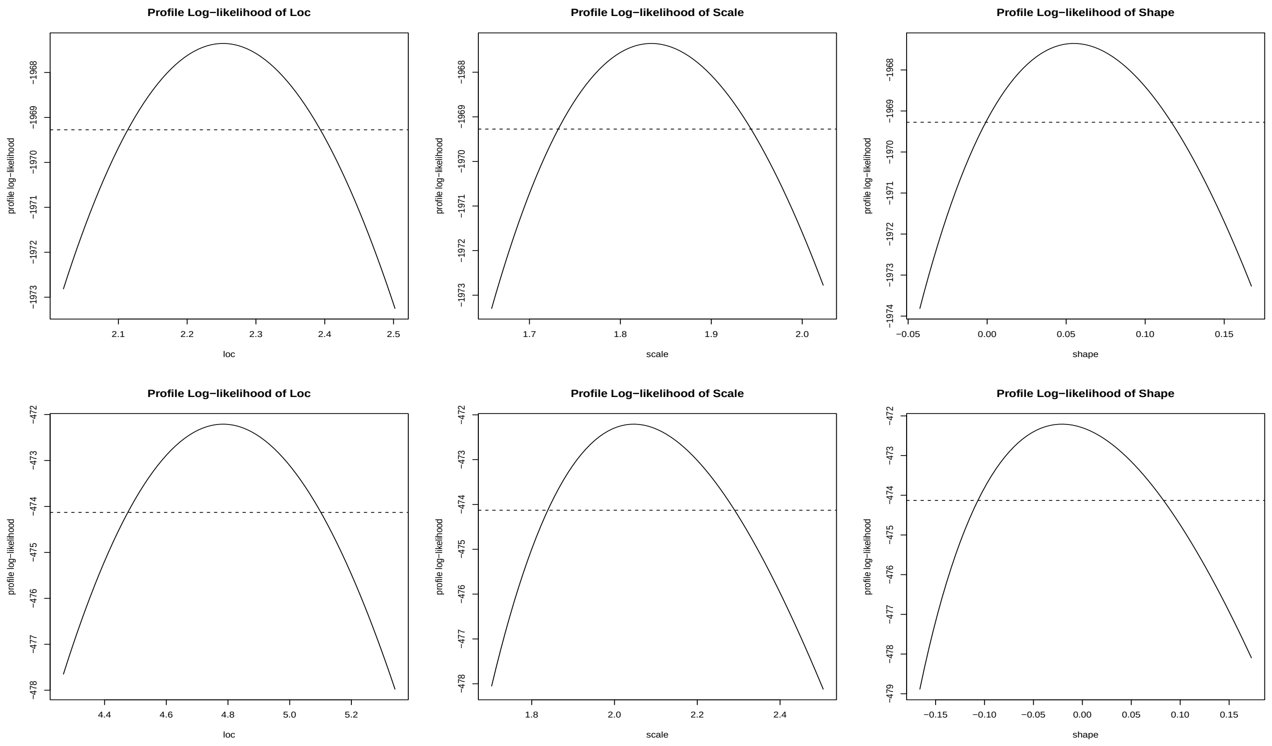

Figure 5: Profile log likelihoods for the three GEV parameters. The top panel corresponds to weekly maxima and the bottom panel to monthly maxima.

```
# here we take a fixed threshold u=5, but you can choose u based on part (a)
(thresh.fit_gpd <- fpot(esk.rain$precip, threshold=5, start=list(scale=1.2,
 shape=0.1)))

 Call: fpot(x = esk.rain$precip, threshold = 5, start = list(scale = 1.2,
  shape = 0.1))
Deviance: 1058.954

Threshold: 5
Number Above: 356
Proportion Above: 0.0024

Estimates
  scale    shape
1.52216  0.06725

Standard Errors
  scale    shape
0.11488  0.05387

# next, we look at the resulting diagnostic plots
par(mfrow=c(1,4))
plot(thresh.fit_gpd)
```

The diagnostic plots in Figure 8 show that the GPD fits the data quite well.

Solution 3

(a) Running the code below gives Figure 9.

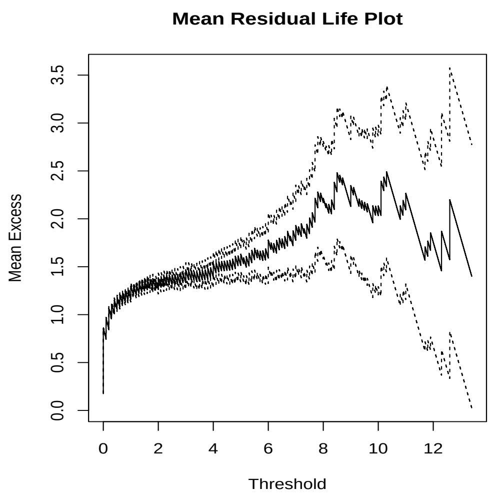

Figure 6: Mean residual life plot

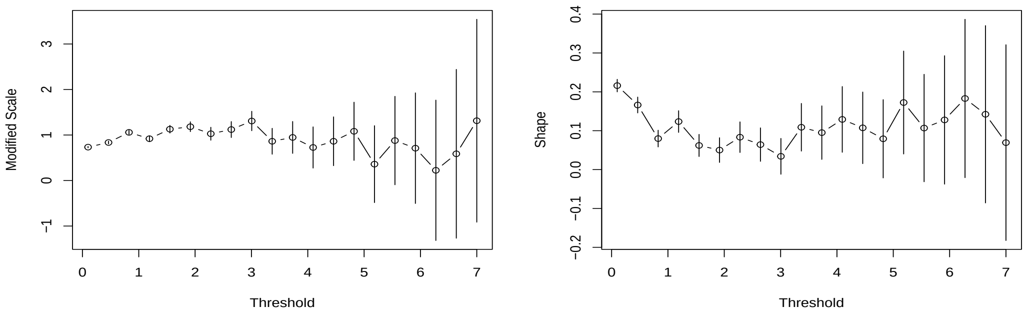

Figure 7: Parameter stability plots

```
m <- 10
fit.w <- fgev(weekly.max, prob=1/(m*52)) # fit to weekly maxima
fit.m <- fgev(monthly.max, prob=1/(m*12)) # fit to monthly maxima
# compare profile log likelihoods for the two fits
par(mfrow=c(2,3))
plot(profile(fit.w))
plot(profile(fit.m))
```

The profile log-likelihood plots show certain differences between the return levels and their confidence intervals, with higher return levels estimated from the weekly maxima. In addition, the confidence intervals for the return levels are asymmetric, in particular for monthly maxima, so one should not compute symmetric confidence intervals based directly on the MLE (e.g., slide 24). For a closer inspection, we look at the estimates using the code

```
fit.w

Call: fgev(x = weekly.max, prob = 1/(m * 52))
Deviance: 3934.709

Estimates
quantile     scale     shape
15.92728   1.83354   0.05482
```

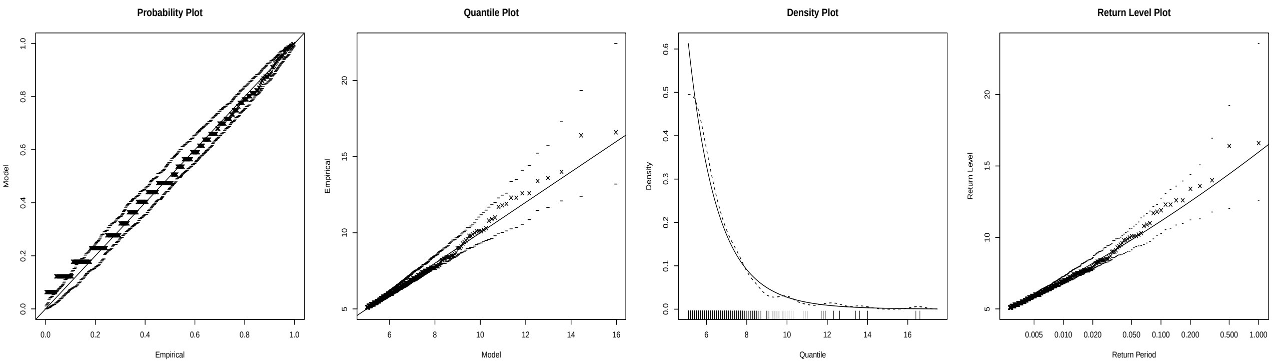

Figure 8: Diagnostic plots for the GPD fit

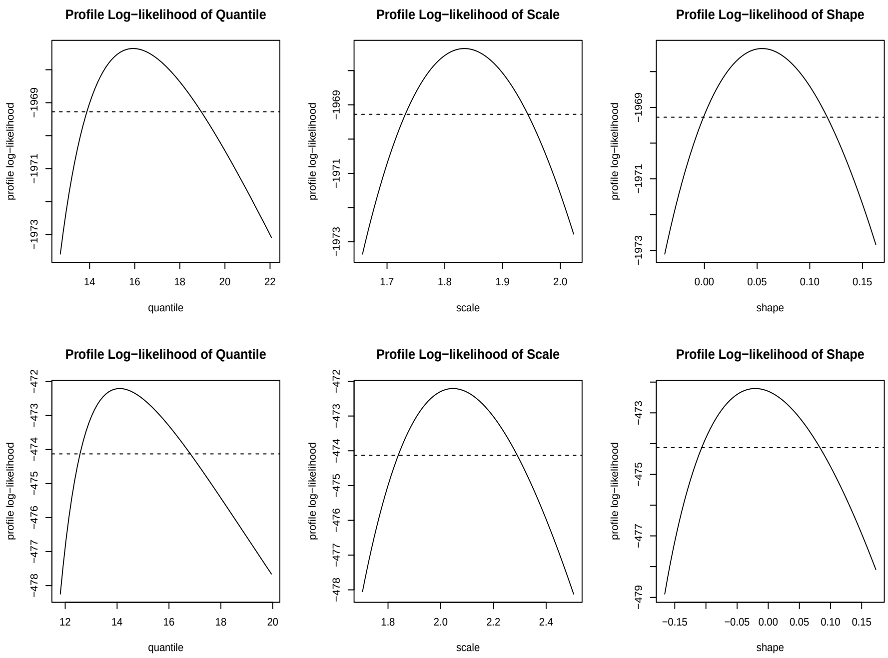

Figure 9: Profile log-likelihood plots for weekly maxima (top panel) and monthly maxima (bottom panel) of precipitation.

```
Standard Errors
quantile     scale     shape
 1.29123   0.05415   0.03082

Optimization Information
  Convergence: successful
  Function Evaluations: 49
  Gradient Evaluations: 13

fit.m

Call: fgev(x = monthly.max, prob = 1/(m * 12))
Deviance: 944.4168

Estimates
quantile     scale     shape
```

```
14.10006   2.04654  -0.02089

Standard Errors
quantile     scale     shape
 1.01697   0.11445   0.04842
```

The scale and shape parameter are very close to the MLEs from Problem 1, shown in Table 3.

	$\widehat{\mu}_{\text{MLE}}$	$\widehat{\sigma}_{\text{MLE}}$	$\widehat{\xi}_{\text{MLE}}$
Weekly	2.25 (0.07)	1.83 (0.05)	0.05 (0.03)
Monthly	4.78 (0.16)	2.05 (0.11)	-0.02 (0.05)

Table 3: Parameter estimates of the fitted GEV models from Problem 1.

We can check whether the return levels estimated above correspond to those computed using the parameters in Table 3; recall from slide 83 that we can compute return levels from the parameters of the GEV via the formula

$$x_{p,c} = \mu_c + \tfrac{\sigma_c}{\xi_c}\left[\{-m_c \log(1-p_c)\}^{-\xi_c} - 1\right], \tag{1}$$

where we use the subscript c to denote the choice of time period of the maxima used for estimation, i.e., weekly or monthly.

Here one has to pay attention to the number of background observations. For the weekly estimates we apply (1) with $m_w = 24 \times 7 = 168$ hourly observations per week, and take $p_w = 1/(52 \times m \times m_w)$. Similarly, we take $m_m = 24 \times 31$ for the monthly background observations and set $p_m = 1/(12 \times m \times m_m)$. Plugging in the estimates from Table 3 we compute $\widehat{x}_{p,w} = 15.93$ and $\widehat{x}_{p,m} = 14.11$, using the code

```
(c(fit.weekly$estimate[1]+fit.weekly$estimate[2]/fit.weekly$estimate[3]
*(((-24*7*log(1-1/(24*7*52*m)))^-fit.weekly$estimate[3])-1),
fit.monthly$estimate[1]+fit.monthly$estimate[2]/fit.monthly$estimate[3]
*(((-24*31*log(1-1/(24*31*12*m)))^-fit.monthly$estimate[3])-1)))

15.93281    14.10830
```

Both estimates are very close to those from `fit.w` and `fit.m`.

We repeat the same procedure, but now for $m \in \{25, 50\}$ using the code below, which gives Figures 10 and 11.

```
m <- 25
fit.w25 <- fgev(weekly.max, prob=1/(m*52)) # fit to weekly maxima
fit.m25 <- fgev(monthly.max, prob=1/(m*12)) # fit to monthly maxima
# compare profile log likelihoods for the two fits
par(mfrow=c(2,3))
plot(profile(fit.w25))
plot(profile(fit.m25))

m <- 50
fit.w50 <- fgev(weekly.max, prob=1/(m*52)) # fit to weekly maxima
fit.m50 <- fgev(monthly.max, prob=1/(m*12)) # fit to monthly maxima
# compare profile log likelihoods for the two fits
par(mfrow=c(2,3))
plot(profile(fit.w50))
plot(profile(fit.m50))
```

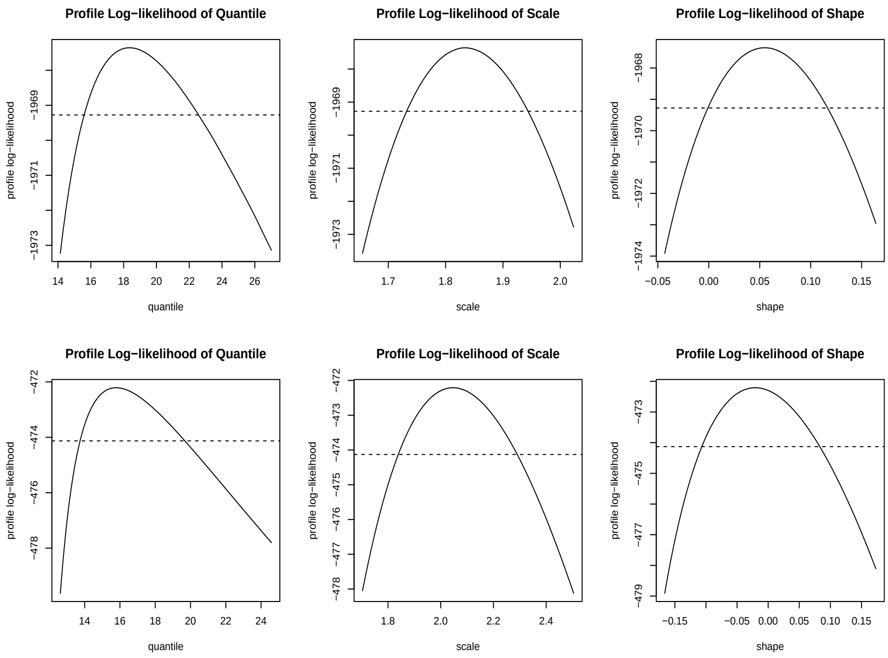

Figure 10: Profile log-likelihood plots for weekly maxima (top panel) and monthly maxima (bottom panel) of precipitation for the 25 year return period.

Both figures show higher return levels, as expected also from (1), since p_c is smaller, but there is a larger increase in the return level from the weekly maxima; we recall that the estimated shape parameter $\widehat{\xi}_w > 0$, whereas $\widehat{\xi}_m < 0$, giving infinite and finite upper bounds, respectively. Therefore, for larger m we would expect $x_{p,w}$ to approach ∞ as $m \to \infty$, whereas $x_{p,m}$ should approach the upper bound $\mu_m - \sigma_m/\xi_m$ (recall Problem 3 from Sheet 4.)

The increase in estimated return levels is accompanied by the increased uncertainty from extrapolating into more extreme regions of the tail. For instance, one can argue by taking into the account the uncertainty of the MLEs in formula (1); if $\xi_w > 0$, we have the term

$$\frac{\widehat{\sigma}_c}{\widehat{\xi}_c}\left[\{-m_w \log(1-p_w)\}^{-\xi_w} - 1\right],$$

where $\left[\{-m_w \log(1-p_w)\}^{-\xi_w} - 1\right] \to \infty$ as $m \to \infty$, and therefore the variance of the estimated $\widehat{x}_{p,w}$ will also increase, thus giving larger confidence intervals.

(b) We now follow the Poisson process approach and use the following code, which gives Figure 12.

```
 u <- 5; m <- 10
 (fit <- fpot(esk.rain$precip, threshold=u, mper=m, npp=365.25*24))
 Call: fpot(x = esk.rain$precip, threshold = u, npp = 365.25 * 24, mper = m)
Deviance: 1058.954

Threshold: 5
Number Above: 356
Proportion Above: 0.0024

Estimates
  rlevel     shape
14.78335   0.06699
```

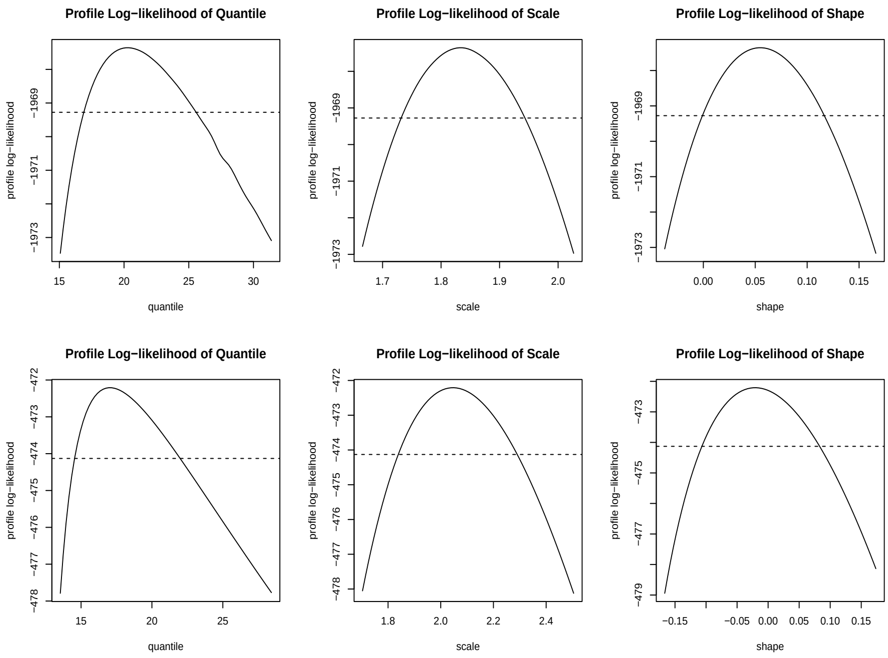

Figure 11: Profile log-likelihood plots for weekly maxima (top panel) and monthly maxima (bottom panel) of precipitation for the 50 year return period.

```
Standard Errors
 rlevel    shape
1.14293  0.05382

 par(mfrow=c(1,2))
 plot(profile(fit))
```

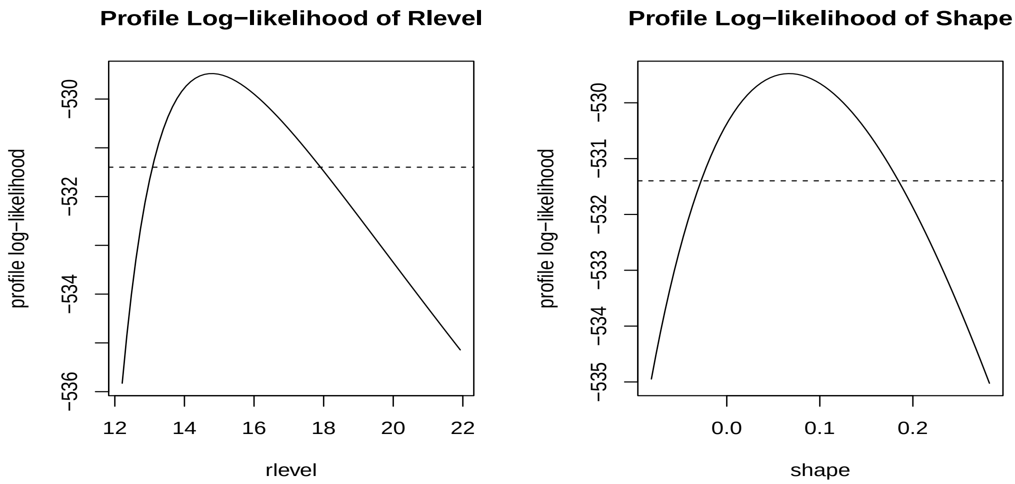

Figure 12: Profile log-likelihood of the return level (left) and shape parameter (right) for the Poisson process.

We also run the following code and, to compare the results, illustrate in Figure 13 only the profile log-likelihood plots of the return levels for different periods m.

```
m <- 25
```

```
(fit25 <- fpot(esk.rain$precip, threshold=u, mper=m, npp=365.25*24))

m <- 50
(fit50 <- fpot(esk.rain$precip, threshold=u, mper=m, npp=365.25*24))
```

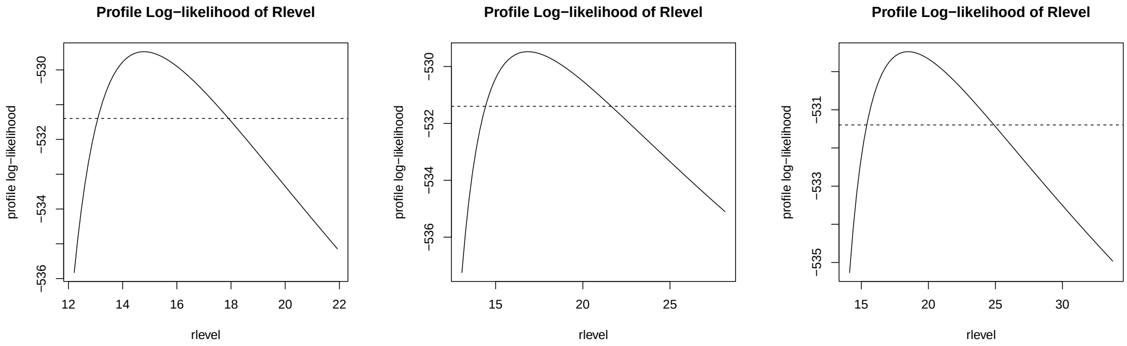

Figure 13: Profile log-likelihood plots for 10 (left), 25 (center) and 50 (right) year return level for the Poisson process.

Figures 9 and 12 show a smaller difference between the estimated return levels from the Poisson process and the monthly maxima for $m = 10$; however, note that the estimated shape parameter under the Poisson process approach is $\widehat{\xi}_{PP} = 0.07$, whereas it is negative for `fit.m`. The confidence intervals are also similar. However, as we consider a longer return period, we notice that both the return level and the confidence intervals are larger for the Poisson process approach relative to the fit from the monthly maxima in (a).

We have from the reality-check that the hourly maximum over 17 years is 15 mm, so the return level $\widehat{x}_{PP} = 14.8$ mm for $m = 10$ is a fairly reasonable estimate.

We now let the threshold u vary using the code below. We observe in Figure 14 differences in the estimated return levels and the shape parameters, as well as the resulting confidence intervals due to the additional or fewer exceedances, an example of bias-variance tradeoff.

```
u <- 2; m <- 10
(fitu2 <- fpot(esk.rain$precip, threshold=u, mper=m, npp=365.25*24))

Call: fpot(x = esk.rain$precip, threshold = u, npp = 365.25 * 24, mper = m)
Deviance: 9080.691

Threshold: 2
Number Above: 3426
Proportion Above: 0.023

Estimates
  rlevel     shape
13.47979   0.02993

Standard Errors
 rlevel    shape
0.58314  0.01597

par(mfrow=c(1,2))
plot(profile(fitu2))
```

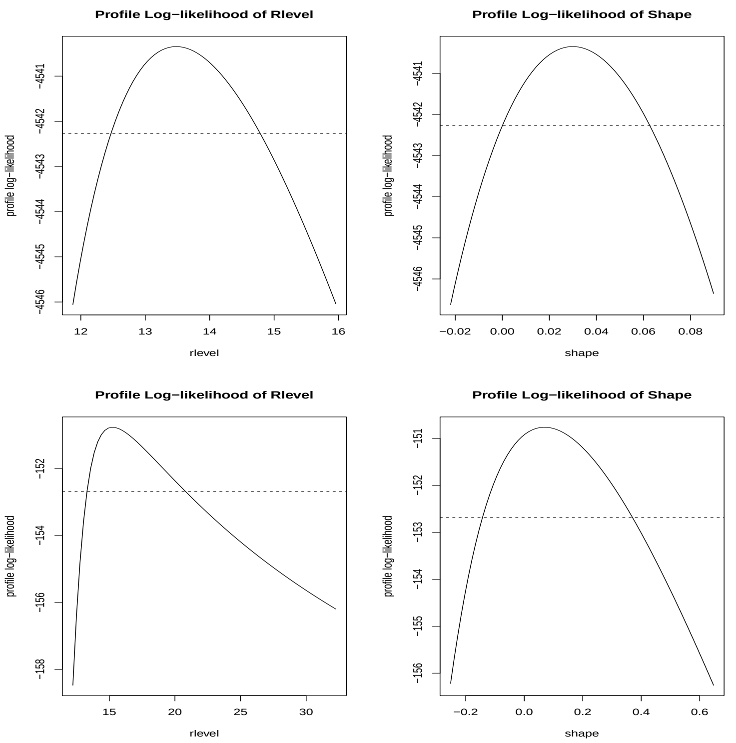

Figure 14: Profile log-likelihood plots for $u = 2$ (top) and $u = 7$ (bottom) for the Poisson process.

```
u <- 7; m <- 10
(fitu7 <- fpot(esk.rain$precip, threshold=u, mper=m, npp=365.25*24))

Call: fpot(x = esk.rain$precip, threshold = u, npp = 365.25 * 24, mper = m)
Deviance: 301.5227

Threshold: 7
Number Above: 91
Proportion Above: 6e-04

Estimates
  rlevel     shape
15.24709   0.06932

Standard Errors
rlevel   shape
1.5124  0.1285

par(mfrow=c(1,2))
plot(profile(fitu7))
```