An Introduction to Statistical Modelling of Extreme Values

```
library(ismev)

## Loading required package: mgcv

## Loading required package: nlme

## This is mgcv 1.8-23. For overview type 'help("mgcv-package")'.

library(extRemes)

## Loading required package: Lmoments

## Loading required package: distillery

## Loading required package: car

##

## Attaching package: 'extRemes'

## The following objects are masked from 'package:stats':

##

## qqnorm, qqplot

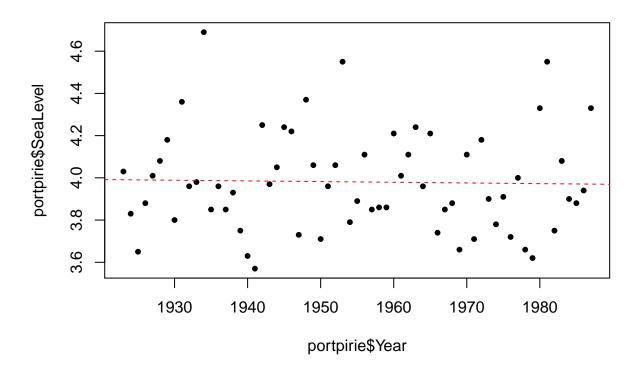
source("extreme_functions.r")
```

Chapter 6: Non-stationary squence

Port Pirie

```
library(ismev)
data("portpirie")

plot(x=portpirie$Year, y=portpirie$SeaLevel, pch=20)
fit <- lm(SeaLevel ~ Year, data = portpirie)
abline(fit, col="red", lty=2)</pre>
```



summary(fit)

```
##
## Call:
## lm(formula = SeaLevel ~ Year, data = portpirie)
##
## Residuals:
##
       Min
                1Q Median
                                3Q
                                       Max
  -0.4151 -0.1606 -0.0280 0.1316 0.7026
##
  Coefficients:
##
##
                 Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.6086420 3.1322646
                                                0.146
                                       1.471
## Year
               -0.0003212 0.0016021 -0.201
                                                0.842
##
\#\# Residual standard error: 0.2423 on 63 degrees of freedom
## Multiple R-squared: 0.0006378, Adjusted R-squared:
## F-statistic: 0.0402 on 1 and 63 DF, p-value: 0.8417
```

Fit with normal model

```
fitgev <- gev.fit(portpirie$SeaLevel)
## $conv
## [1] 0</pre>
```

```
##
## $nllh
## [1] -4.339058
##
## $mle
## [1] 3.87474692 0.19804120 -0.05008773
##
## $se
## [1] 0.02793211 0.02024610 0.09825633
```

Fit with linear time model for u

```
ti <- matrix(ncol=1,nrow=length(portpirie$SeaLevel))</pre>
ti[,1] \leftarrow seq(1,65,1)
fitgev.ut <- gev.fit(portpirie$SeaLevel, ydat = ti, mul = 1)</pre>
## $model
## $model[[1]]
## [1] 1
##
## $model[[2]]
## NULL
##
## $model[[3]]
## NULL
##
##
## [1] "c(identity, identity, identity)"
##
## $conv
## [1] 0
##
## $nllh
## [1] -4.375107
##
## $mle
## [1] 3.8865633240 -0.0003548552 0.1979732663 -0.0504552135
##
## $se
## [1] 0.051248878 0.001306864 0.020207615 0.097786572
```

Ratio test

No difference

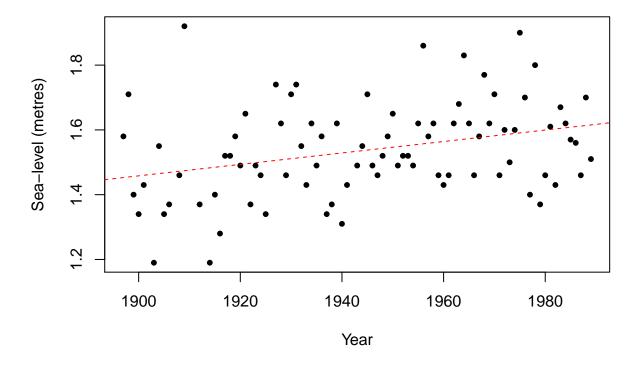
```
2*(-fitgev.ut$nllh - -fitgev$nllh)
## [1] 0.07209759
```

Sea levels ans Southern Oscillation Index

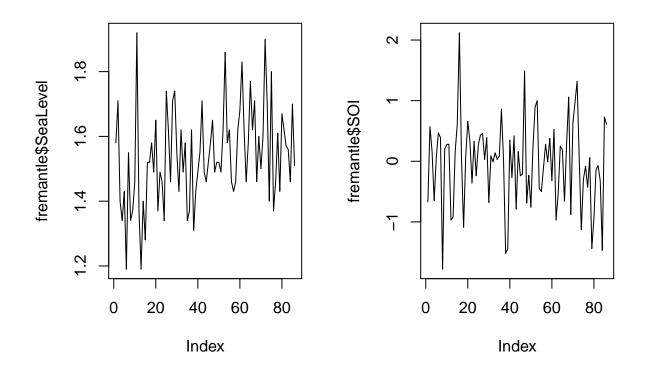
Data

The fremantle data frame has 86 rows and 3 columns. The second column gives 86 annual maximimum sea levels recorded at Fremantle, Western Australia, within the period 1897 to 1989. The first column gives the corresponding years. The third column gives annual mean values of the Southern Oscillation Index (SOI), which is a proxy for meteorological volitility.

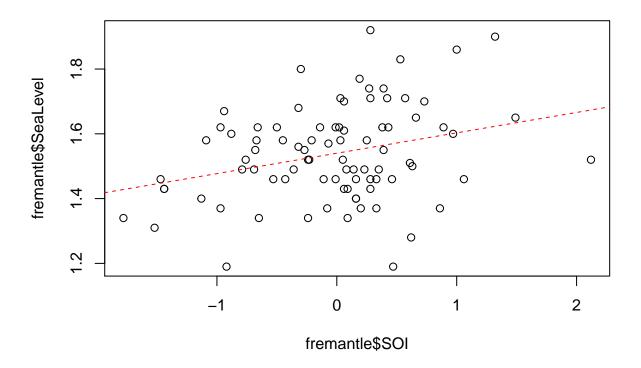
```
data("fremantle")
head(fremantle)
##
     Year SeaLevel
                     SOI
## 1 1897
              1.58 -0.67
## 2 1898
              1.71 0.57
## 3 1899
              1.40 0.16
## 4 1900
              1.34 -0.65
## 5 1901
              1.43 0.06
## 7 1903
              1.19
                    0.47
plot(x=fremantle$Year, fremantle$SeaLevel, pch=20,
       ylab="Sea-level (metres)", xlab="Year")
fit <- lm(SeaLevel ~ Year, data = fremantle)</pre>
abline(fit, col="red", lty=2)
```



```
par(mfrow=c(1,2))
plot(fremantle$SeaLevel, t="1")
plot(fremantle$SOI, t="1")
```



```
par(mfrow=c(1,1))
plot(fremantle$SOI, fremantle$SeaLevel)
abline(lm(SeaLevel ~ SOI, data = fremantle), col="red", lty=2)
```



Covariates

```
n <- dim(fremantle)[1]</pre>
#Covariates
covar <- matrix(ncol = 3, nrow = n)</pre>
covar[,1] <- fremantle$SOI</pre>
covar[,2] <- seq(1, n, 1) # Linear trend</pre>
covar[,3] <- covar[,2]^2 # Quadratic trend</pre>
head(covar)
          [,1] [,2] [,3]
## [1,] -0.67
                        1
## [2,]
         0.57
## [3,]
         0.16
                        9
## [4,] -0.65
                       16
## [5,]
         0.06
                       25
## [6,]
         0.47
                       36
Model stationary
fitgev <- gev.fit(fremantle$SeaLevel, show = F)</pre>
fitgev$nllh
```

```
## [1] -43.56663
fitgev$mle
## [1] 1.4823409 0.1412671 -0.2174320
fitgev$se
## [1] 0.01672502 0.01149461 0.06377394
Model with SOI
fitgev.SOI <- gev.fit(fremantle$SeaLevel, ydat = covar, mul = 1, show = F)</pre>
fitgev.SOI$nllh
## [1] -47.21114
fitgev.SOI$mle
## [1] 1.48985338 0.06188902 0.13960518 -0.26848380
fitgev.SOI$se
## [1] 0.01655406 0.02315637 0.01150991 0.06399288
Model with Linear time
fitgev.ut <- gev.fit(fremantle$SeaLevel, ydat = covar, mul = 2, show = F)</pre>
fitgev.ut$nllh
## [1] -49.78972
fitgev.ut$mle
## [1] 1.387186155 0.002140832 0.124716473 -0.128545018
fitgev.ut$se
## [1] 0.0274796482 0.0005215259 0.0104146285 0.0679844086
Model with quadratic time
fitgev.ut2 <- gev.fit(fremantle$SeaLevel, ydat = covar, mul = c(2,3), show = F)</pre>
## Warning in sqrt(diag(z$cov)): NaNs produced
fitgev.ut2$nllh
## [1] -50.95252
fitgev.ut2$mle
## [1] 1.331932e+00 5.642570e-03 -3.921111e-05 1.208444e-01 -9.821101e-02
fitgev.ut2$se
```

NaN 0.0090983484 0.0060969134

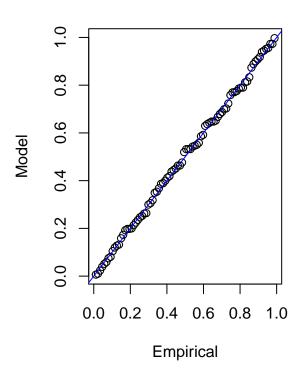
[1] 0.0265635286 0.0005200551

Model with SOI and Linear time

```
fitgev.SOIut <- gev.fit(fremantle$SeaLevel, ydat = covar, mul = c(1, 2), show = F)
fitgev.SOIut$nllh
## [1] -53.8257
fitgev.SOIut$mle
## [1] 1.389381297 0.055171074 0.002232467 0.121147089 -0.154480161
fitgev.SOIut$se
## [1] 0.0272538644 0.0197789753 0.0005178779 0.0100390306 0.0636920071
Model with linear time for sigma
fitgev.st <- gev.fit(fremantle$SeaLevel, ydat = covar, sigl = 2, show = F)</pre>
fitgev.st$nllh
## [1] -44.67998
fitgev.st$mle
## [1] 1.4920666018 0.1666643314 -0.0007118848 -0.1609604077
fitgev.st$se
## [1] 1.587629e-02 1.122782e-02 1.999940e-06 7.249364e-02
Ratio test between stationary and linear trend
2*(-fitgev.ut$nllh - -fitgev$nllh)
## [1] 12.44618
Diag
gev.diag(fitgev.ut)
```

Residual Probability Plot

Residual Quantile Plot (Gumbel Sc



3 1933 121 113 106 105 102 89 89 88 86

91

90

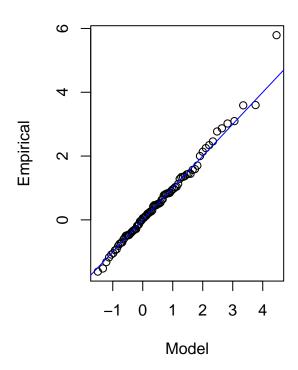
91 89 88 88

93 91 NA NA NA

87 87 87 84 82

91

93



Sea level in Venice

Data

The venice data frame has 51 rows and 11 columns. The final ten columns contain the 10 largest sea levels observed within the year given by the first column. The ten largest sea levels are given for every year in the period 1931 to 1981, excluding 1935 in which only the six largest measurements are available. SeaLevai is measured in cm.

```
library(ismev)
data("venice")
head(venice)
                   r3
     Year
          r1
               r2
                       r4
                            r5 r6 r7 r8 r9 r10
## 1 1931 103
               99
                            94 89 86 85 84
                   98
                        96
## 2 1932
          78
               78
                   74
                       73
                            73 72 71 70 70
                                            69
```

85

86

Covariates

4 1934 116 113

6 1936 147 106

5 1935 115 107 105 101

```
n <- dim(venice)[1]</pre>
#Covariates
covar <- matrix(ncol = 1, nrow = n)</pre>
covar[,1] <- seq(1, n, 1) # Linear trend</pre>
head(covar)
##
         [,1]
## [1,]
## [2,]
            2
## [3,]
            3
## [4,]
## [5,]
          5
## [6,]
\mathbf{Fit}
r=1 \text{ model}
venice.r1 <- venice$r1</pre>
fitgev.r1 <- gev.fit(venice.r1, show = F)</pre>
fitgev.r1.ut <- gev.fit(venice.r1, ydat = covar, mul = 1, show = F)</pre>
r=5 model
venice.r5 <- venice[,c(2:6)]</pre>
fitgev.r5 <- rlarg.fit(venice.r5, show = F)</pre>
fitgev.r5.ut <- rlarg.fit(venice.r5, ydat = covar, mul = 1, show = F)</pre>
r=10 model
venice.r10 <- venice[,-1]</pre>
fitgev.r10 <- rlarg.fit(venice.r10, show = F)</pre>
fitgev.r10.ut <- rlarg.fit(venice.r10, ydat = covar, mul = 1, show = F)</pre>
```

Results for parameters

```
gev.result <- function(gevfit){
  loglik <- -gevfit$nllh
  par.mu0 <- gevfit$mle[1]
  par.mu1 <- NaN
  par.sigma <- gevfit$mle[2]
  par.xi <- gevfit$mle[3]
  se.mu0 <- gevfit$se[1]
  se.mu1 <- NaN
  se.sigma <- gevfit$se[2]
  se.xi <- gevfit$se[3]

result <- data.frame(loglik=loglik, mu0=par.mu0, se.mu0=se.mu0, mu1=par.mu1, se.mu1=se.mu1,</pre>
```

```
sigma=par.sigma, se.sigma=se.sigma,
                          xi=par.xi, se.xi=se.xi)
  return(round(result, 3))
}
gev.ut.result <- function(gevfit){</pre>
  loglik <- -gevfit$nllh</pre>
  par.mu0 <- gevfit$mle[1]</pre>
  par.mu1 <- gevfit$mle[2]</pre>
  par.sigma <- gevfit$mle[3]</pre>
  par.xi <- gevfit$mle[4]</pre>
  se.mu0 <- gevfit$se[1]</pre>
  se.mu1 <- gevfit$se[2]</pre>
  se.sigma <- gevfit$se[3]</pre>
  se.xi <- gevfit$se[4]</pre>
  result <- data.frame(loglik=loglik, mu0=par.mu0, se.mu0=se.mu0,
                          mu1=par.mu1, se.mu1=se.mu1,
                          sigma=par.sigma, se.sigma=se.sigma,
                          xi=par.xi, se.xi=se.xi)
  return(round(result, 3))
```

Stationary models

```
rbind(gev.result(fitgev.r1), gev.result(fitgev.r5),gev.result(fitgev.r10))

## loglik mu0 se.mu0 mu1 se.mu1 sigma se.sigma xi se.xi
## 1 -222.715 111.099 2.628 NaN NaN 17.175 1.803 -0.077 0.074

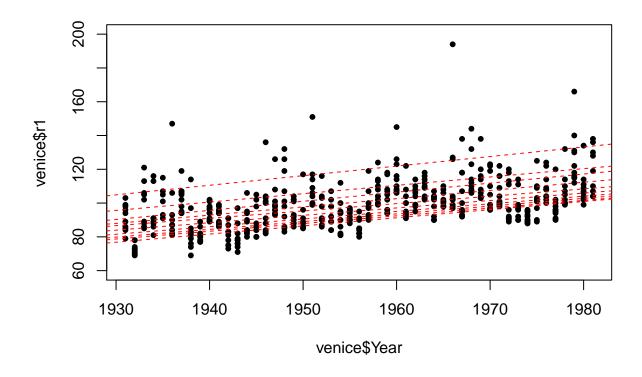
## 2 -731.967 118.569 1.567 NaN NaN 13.662 0.776 -0.088 0.033
## 3 -1139.090 120.548 1.362 NaN NaN 12.784 0.549 -0.113 0.020
```

Non-stationary models

2 -704.760 104.233 2.038 0.458 0.055 12.290 0.805 -0.037 0.042 ## 3 -1084.059 104.513 1.667 0.482 0.041 11.737 0.641 -0.065 0.028

Plot with linear lines

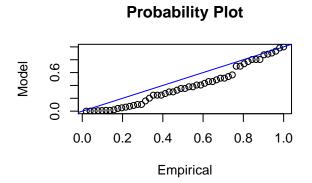
```
plot(x=venice$Year, y=venice$r1, pch=20, ylim=c(60,200))
for(i in 3:11){
   points(x=venice$Year, y=venice[,i], pch=20)
}
for(i in 2:11){
   abline(lm(venice[,i] ~ venice$Year), col="red", lty=2)
}
```

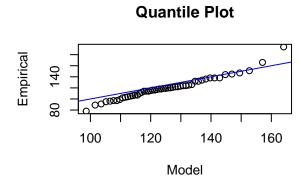


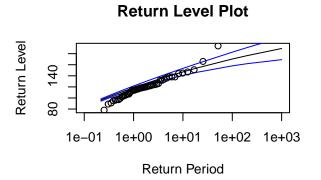
Models check

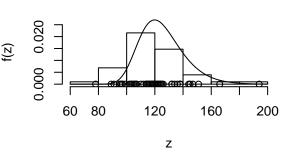
Stationary model

rlarg.diag(fitgev.r5)

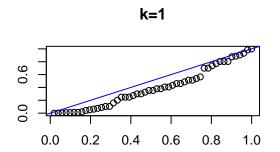


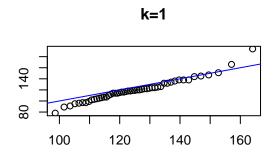


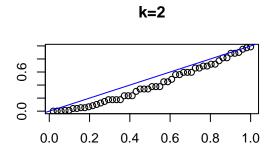


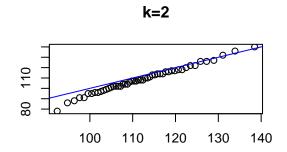


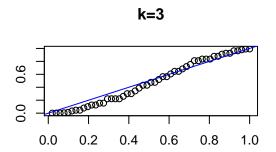
Density Plot

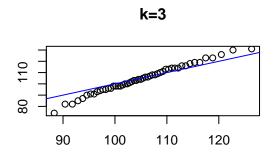


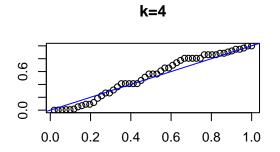


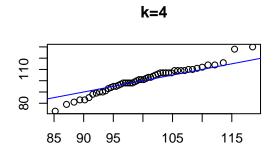


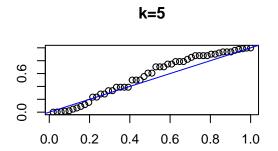


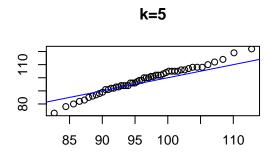




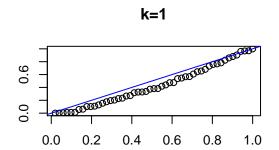


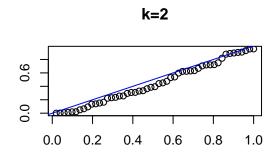


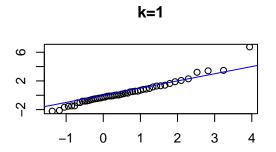


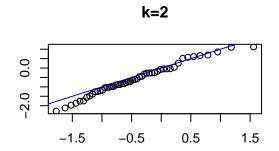


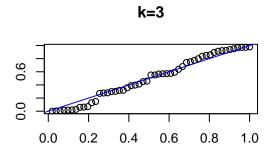
rlarg.diag(fitgev.r5.ut)

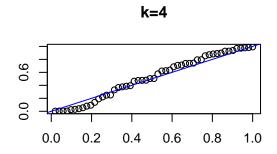


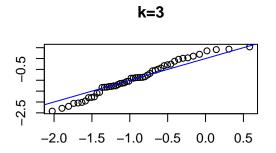


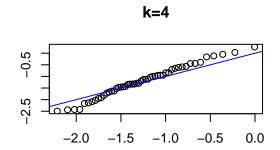


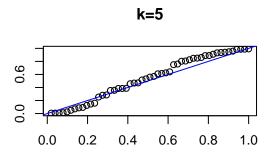


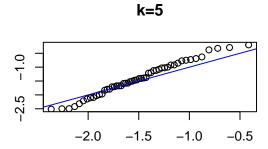












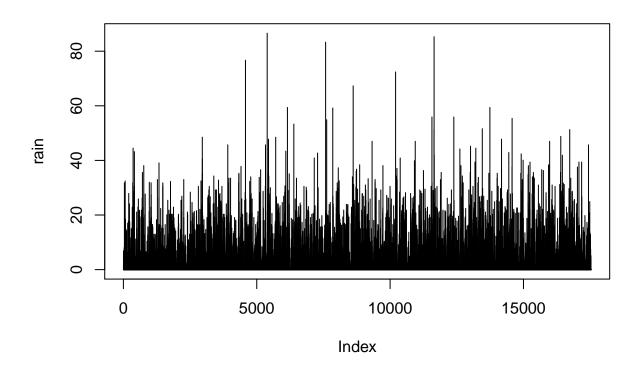
Daily rainfall

Data

A numeric vector containing daily rainfall accumulations at a location in south-west England over the period 1914 to 1962.

```
library(ismev)
data("rain")
head(rain)

## [1] 0.0 2.3 1.3 6.9 4.6 0.0
plot(rain, t="h")
```



Covariates

```
n <- length(rain)[1]

#Covariates
covar <- matrix(ncol = 1, nrow = n)
covar[,1] <- seq(1, n, 1) # Linear trend

head(covar,3)

## [,1]
## [1,] 1
## [2,] 2
## [3,] 3</pre>
```

Fit with stationary model

```
fitgpd <- gpd.fit(rain, threshold = 30, show = F)
fitgpd$nllh

## [1] 485.0937
fitgpd$mle

## [1] 7.4422639 0.1843027</pre>
```

Fit with non-stationary model

```
fitgpd.st <- gpd.fit(rain, 30, ydat = covar, sigl = 1, siglink = exp, show = F)

## Warning in sqrt(diag(z$cov)): NaNs produced

fitgpd.st$nllh

## [1] 484.6016

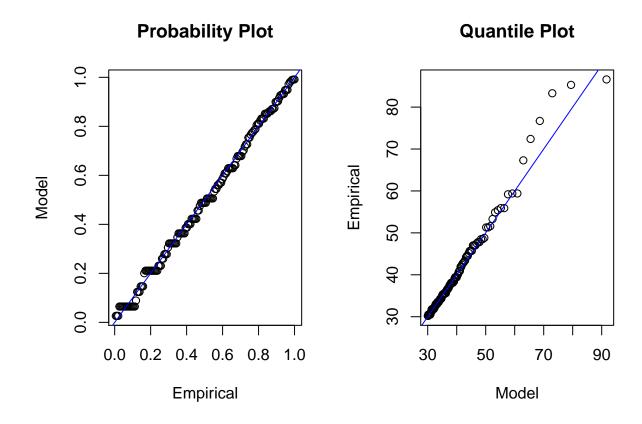
fitgpd.st$mle</pre>
```

[1] 1.8039221883 0.0000196265 0.1977440505

Model check

Stationary model

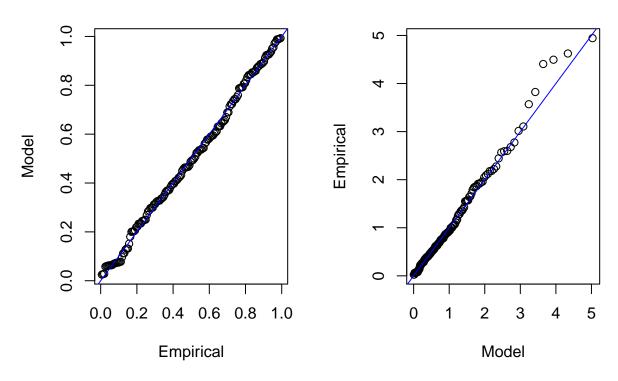
```
par(mfrow=c(1,2))
cgpd.pp(fitgpd); cgpd.qq(fitgpd)
```



gpd.diag(fitgpd.st)

Residual Probability Plot

Residual Quantile Plot (Exptl. Sca



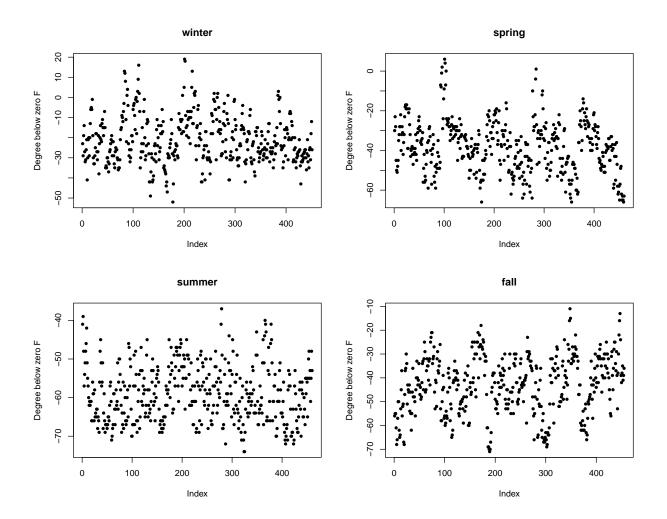
Wooster temperature

Data

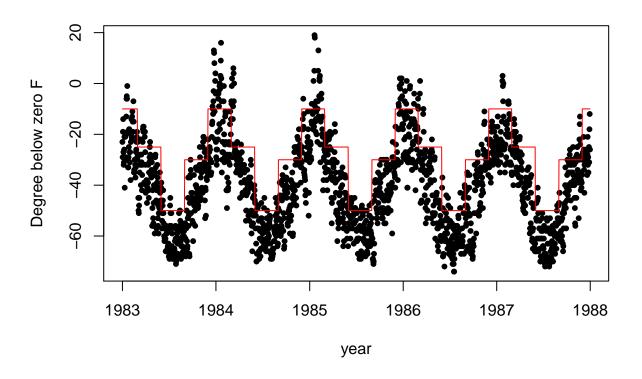
```
data("wooster")
wooster.dat <- data.frame(temp=-wooster)</pre>
start.date <- as.Date("01-01-1983", format="%d-%m-%Y")
end.date <- as.Date("01-01-1988", format="%d-%m-%Y")
dates <- seq(start.date, end.date, by=1)</pre>
dates <- dates[-length(dates)]</pre>
wooster.dat <- cbind(wooster.dat, dates)</pre>
wooster.dat$year <- as.numeric(format(as.Date(wooster.dat$dates,</pre>
                                                  format="%d/%m/%Y"),"%Y"))
wooster.dat$month <- as.numeric(format(as.Date(wooster.dat$dates,</pre>
                                                   format="%d/%m/%Y"), "%m"))
wooster.dat$spring[wooster.dat$month==3 |
                      wooster.dat$month==4 |
                       wooster.dat$month==5 ] <- 1</pre>
wooster.dat$spring[is.na(wooster.dat$spring)] <- 0</pre>
wooster.dat$summer[wooster.dat$month==6 |
                      wooster.dat$month==7 |
```

```
wooster.dat$month==8 ] <- 1</pre>
wooster.dat$summer[is.na(wooster.dat$summer)] <- 0</pre>
wooster.dat$fall[wooster.dat$month==9 |
                   wooster.dat$month==10 |
                   wooster.dat$month==11 ] <- 1</pre>
wooster.dat$fall[is.na(wooster.dat$fall)] <- 0</pre>
wooster.dat$winter[wooster.dat$month==12 |
                     wooster.dat$month==1 |
                   wooster.dat$month==2 ] <- 1</pre>
wooster.dat$winter[is.na(wooster.dat$winter)] <- 0</pre>
head(wooster.dat)
               dates year month spring summer fall winter
##
     temp
## 1 -23 1983-01-01 1983
## 2 -29 1983-01-02 1983
                                                  0
                                     0
                                             0
                                                         1
## 3 -19 1983-01-03 1983
                              1
                                     0
                                                  0
## 4 -14 1983-01-04 1983
                             1
                                    0
                                                 0
## 5 -27 1983-01-05 1983
                                                 0
                             1
                           1
## 6 -32 1983-01-06 1983
                                 0
```

Plot by season



Seasonal threshold



Fit with stationary model

[1] 1.48317853 0.09270039

```
fitgpd21 <- gpd.fit(wooster.dat$temp, threshold = -10)</pre>
## $threshold
## [1] -10
##
## $nexc
## [1] 85
##
## $conv
##
   [1] 0
##
## $nllh
## [1] 262.1725
##
## $mle
## [1] 10.6402671 -0.2802218
##
## $rate
## [1] 0.04654984
##
```

Non-stationary model

```
fitgpd2 <- gpd.fit(wooster.dat$temp, threshold = wooster.dat$thres)</pre>
## $model
## $model[[1]]
## NULL
##
## $model[[2]]
## NULL
##
##
## $link
## [1] "c(identity, identity)"
##
## $nexc
## [1] 235
## $conv
## [1] 0
##
## $nllh
## [1] 684.7317
##
## $mle
## [1] 7.41154486 -0.08939142
## $rate
## [1] 0.1286966
##
## $se
## [1] 0.6664141 0.0621433
```

Sea level in Venice (UNC)

Data

```
library(ismev)
data("venice")
head(venice)

## Year r1 r2 r3 r4 r5 r6 r7 r8 r9 r10

## 1 1931 103 99 98 96 94 89 86 85 84 79

## 2 1932 78 78 74 73 73 72 71 70 70 69

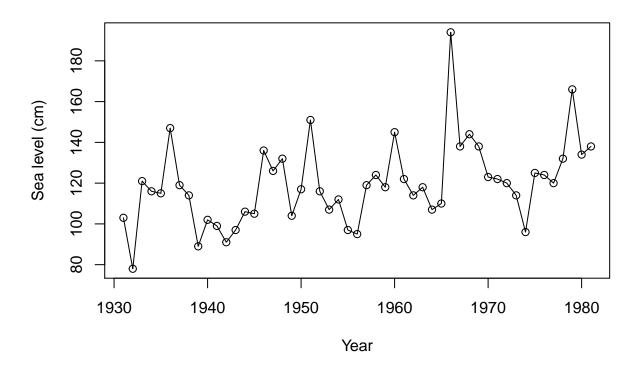
## 3 1933 121 113 106 105 102 89 89 88 86 85

## 4 1934 116 113 91 91 91 89 88 88 86 81

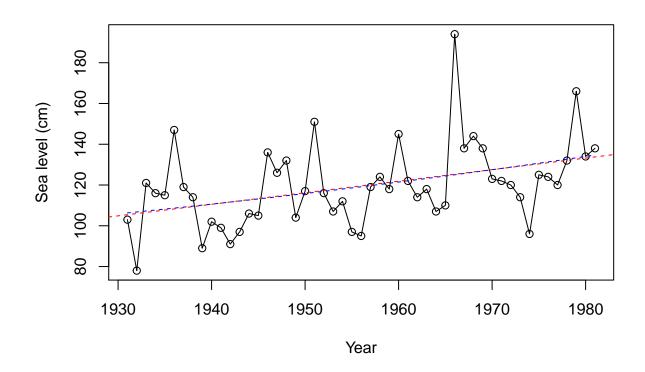
## 5 1935 115 107 105 101 93 91 NA NA NA

## 6 1936 147 106 93 90 87 87 87 84 82 81
```

Maxima



Trend analysis



summary(fit)

```
##
## Call:
## lm(formula = r1 ~ Year, data = venice.anmax)
##
## Residuals:
       Min
                1Q Median
##
                                3Q
                                      Max
## -33.813 -11.211 -3.309
                             9.515
                                  68.722
##
## Coefficients:
##
                Estimate Std. Error t value Pr(>|t|)
                           346.4770 -2.856 0.00628 **
## (Intercept) -989.3822
                             0.1771
                                     3.201 0.00241 **
## Year
                  0.5670
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 18.62 on 49 degrees of freedom
## Multiple R-squared: 0.1729, Adjusted R-squared: 0.1561
## F-statistic: 10.25 on 1 and 49 DF, p-value: 0.002406
summary(fit2)
##
## Call:
## lm(formula = r1 ~ poly(Year, 2, raw = TRUE), data = venice.anmax)
##
```

```
## Residuals:
##
      Min
               1Q Median 3Q
                                      Max
## -34.047 -10.865 -3.322 9.067 68.976
##
## Coefficients:
##
                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)
                              7.334e+03 5.202e+04 0.141
## poly(Year, 2, raw = TRUE)1 -7.944e+00 5.319e+01 -0.149
                                                             0.882
## poly(Year, 2, raw = TRUE)2 2.176e-03 1.360e-02 0.160
                                                             0.874
## Residual standard error: 18.81 on 48 degrees of freedom
## Multiple R-squared: 0.1734, Adjusted R-squared: 0.1389
## F-statistic: 5.034 on 2 and 48 DF, p-value: 0.01036
GEV
fitgev <- gev.fit(venice.anmax$r1)</pre>
```

Stationary

```
## $conv
## [1] 0
##
## $nllh
## [1] 222.7145
##
## $mle
## [1] 111.09925486 17.17548761 -0.07673265
##
## $se
## [1] 2.6280070 1.8033672 0.0735214
```

Linear trend for location

```
#Time index
ti <- matrix(ncol=1,nrow=length(venice.anmax$Year))</pre>
ti[,1] <- seq(1,length(venice.anmax$Year),1)</pre>
#Fit
fitgev.ut <- gev.fit(venice.anmax$r1, ydat = ti, mul = 1)</pre>
## $model
## $model[[1]]
## [1] 1
##
## $model[[2]]
## NULL
##
## $model[[3]]
## NULL
##
## $link
## [1] "c(identity, identity, identity)"
```

```
##
## $conv
## [1] 0
##
## $nllh
## [1] 216.0626
##
## $mle
## [1] 96.98579330  0.56414269 14.58435088 -0.02731421
##
## $se
## [1] 4.24930969 0.13948421 1.57840034 0.08270996
```

```
Quandratic trend for location
#Time index
ti2 <- matrix(ncol = 2, nrow = length(venice.anmax$Year))</pre>
ti2[,1] <- seq(1,length(venice.anmax$Year),1)</pre>
ti2[,2] <- (ti2[,1])^2
fitgev.ut2 <- gev.fit(venice.anmax$r1, ydat = ti2, mul = c(1,2))</pre>
## $model
## $model[[1]]
## [1] 1 2
## $model[[2]]
## NULL
##
## $model[[3]]
## NULL
##
##
## $link
## [1] "c(identity, identity, identity)"
## $conv
## [1] 0
##
## $nllh
## [1] 216.0555
##
##
## [1] 6.64848069 0.59400203 0.01130770 1.59752694 0.08578017
```

Model selection

Linear and Stationary

```
2*(-fitgev.ut$nllh - -fitgev$nllh)

## [1] 13.30386
as.numeric(pchisq(-fitgev.ut$nllh - -fitgev$nllh, df=1, lower.tail=FALSE))

## [1] 0.009904848
fitgev.ut is better
```

Linear and Quadratic

```
2*(-fitgev.ut2$nllh - -fitgev.ut$nllh)
## [1] 0.01417277
as.numeric(pchisq(-fitgev.ut2$nllh - -fitgev.ut$nllh, df=1, lower.tail=FALSE))
## [1] 0.9329128
```

No difference -> choose linear model

Model diagnostics

gev.diag(fitgev.ut)

Residual Probability Plot

Residual Quantile Plot (Gumbel Sc

