

**PROBLEM APPLICATIONS:  
EXERCISE SESSION ON ANALYSES OF EXTREMES**

Rick Katz

Institute for Study of Society and Environment  
National Center for Atmospheric Research  
Boulder, CO USA

email: [rwk@ucar.edu](mailto:rwk@ucar.edu)

Home page: [www.isse.ucar.edu/staff/katz/](http://www.isse.ucar.edu/staff/katz/)

*Sunday, 11 January 2009*

**DESCRIPTION**

This tutorial (including exercises) serves as a companion to the lecture by Richard Smith on Extreme Value Theory. It consists of two parts, the first focuses on climate applications of extreme value theory under stationarity, the second on applications under non-stationarity (including both trends and physically-based covariates). The tutorial makes extensive use of the package `extRemes`, GUIs for extreme value analysis available within the open source statistical programming language R.

TUTORIAL: Part 1**INTRODUCTION TO EXTREMES TOOLKIT AND CLIMATE APPLICATIONS**

- (1) Installation of R software
- (2) Installation of Extremes Toolkit
- (3) Climate Applications (Under Stationarity)
  - (3.1) Generalized Extreme Value (GEV) distribution/Block maxima
    - (3.1.1) Simulated data from GEV distribution
    - (3.1.2) Fort Collins annual maximum precipitation [`ftcanmax.R`]
  - (3.2) Generalized Pareto (GP)/Distribution tails  
Phoenix daily minimum temperature [`Tphap.R`]
- (4) *Exercise*
  - Economic damage from hurricanes [`damage.R`]  
(fitting GP distribution to upper tail)

TUTORIAL: Part 2**USE OF EXTREMES TOOLKIT IN CLIMATE CHANGE APPLICATIONS**

(1) Using Covariates in Extremes Toolkit

(2) Climate Change Applications

Block maxima approach:

Phoenix summer minimum temperature (trend) [HEAT.R]

(3) *Exercise*

Port Jervis winter maximum temperature [PORTw.R]

(fitting GEV distribution with Arctic Oscillation as covariate)

**PREPARATION** (if use own laptop)

*Statistical programming language R (open source)*: Download and install on laptop (Prior experience with R not assumed)

*Extremes Toolkit (open source)*: Install `extRemes` package within R (No familiarity with `extRemes` assumed)

**RESOURCES**

R: The R Project for Statistical Computing ([www.r-project.org](http://www.r-project.org))

*Extremes Toolkit*: [www.isse.ucar.edu/extremevalues/evtk.html](http://www.isse.ucar.edu/extremevalues/evtk.html)

*Tutorial*: [www.isse.ucar.edu/extremevalues/tutorial.pdf](http://www.isse.ucar.edu/extremevalues/tutorial.pdf) (pdf) or [www.isse.ucar.edu/extremevalues/tutorial](http://www.isse.ucar.edu/extremevalues/tutorial) (html)

*Statistics of Weather and Climate Extremes*:  
[www.isse.ucar.edu/extremevalues/extreme.html](http://www.isse.ucar.edu/extremevalues/extreme.html)

**REFERENCES**

Coles, S., 2001: *An Introduction to Statistical Modeling of Extreme Values*. Springer, London. ([www.springer.com](http://www.springer.com))

Katz, R.W., M.B. Parlange, and P. Naveau, 2002: "Statistics of extremes in hydrology." *Advances in Water Resources*, **25**, 1287–1304.  
([www.isse.ucar.edu/staff/katz/docs/pdf/awr.pdf](http://www.isse.ucar.edu/staff/katz/docs/pdf/awr.pdf))

Stephenson, A., and E. Gilleland, 2006: "Software for the analysis of extreme events: The current state and future directions." *Extremes*, **8**, 87–109.

## PART 1 OF TUTORIAL ON EXTREMES TOOLKIT

### (1) *Installation of R software*

R-2.8.0-win32.exe

Use default options

### (2) *Installation of Extremes Toolkit*

extRemes (R package for extreme value analysis): extRemes\_1.57.zip

ismev (R package used by extRemes): ismev\_1.3.zip

Within R session (using RGui):

(i) Click on **Packages**

(ii) Select **Install package(s) from local zip files**

Note: Ordinarily, if connected to web, can download/install a package simply by selecting the **Install packages(s)** option instead.

Once packages have been installed, can load Extremes Toolkit by executing R command:

```
library(extRemes)
```

Or from RGui:

(i) Click on **Packages**

(ii) Select **Load package**

(iii) Select **extRemes** from list

Getting started:

To obtain help, execute R command:

```
help(extRemes)
```

If extRemes GUI disappears, execute R command:

```
extremes.gui()
```

Data sets that come with extRemes:

Located at `... \R-2.8.0\library\extRemes\data`

May be convenient to move them (or at least the data sets to be used in tutorial) to:

`My Documents`

Data sets to be used in tutorial:

Ft. Collins annual maximum precipitation: `ftcanmax.R`

Phoenix daily minimum temperature: `Tphap.R`

Hurricane damage: `damage.R`

Phoenix summer minimum temperature: `HEAT.R`

Port Jervis winter maximum temperature (and Arctic Oscillation index): `PORTw.R`

(3) *Climate Applications (Under Stationarity)*

(3.1) *Generalized Extreme Value (GEV) distribution/Block maxima*

(3.1.1) *Simulation from GEV distribution*

(i) *Simulate data*

**File/Simulate Data/Generalized Extreme Value (GEV)**

GEV parameters:

Location parameter ( $\mu$ ): **0**

Scale parameter ( $\sigma$ ): **1**

Shape parameter ( $\xi$ ): **0.2**

Sample size: **100**

Save as: `gevsim.R`

Click on **Generate**

Note: Creates `extRemes` data object `gevsim.R`

(ii) Fit GEV distribution to simulated data

**Analyze/Generalized Extreme Value (GEV) Distribution**

Data Object: `gevsim.R`

Optimization Method: **Nelder-Mead**

Response: `gev.sim`

Plot diagnostics: **check**

Click on **OK**

Note: Fitted output saved in `gev.fit1`

Repeat simulation exercise with Shape parameter ( $\xi$ ) changed to **-0.2**

### (3.1.2) Fort Collins annual maximum precipitation

To obtain more information about this data set, use R command: `help(ftcanmax)`

#### (i) Read Fort Collins data into `extRemes`

##### **File/Read Data**

Select file `ftcanmax.R`

File Type: **R source**

Save as (in R): `ftcanmax.R`

Click on **OK**

Note: Creates `extRemes` data object `ftcanmax.R`

#### (ii) Check data

##### **Plot/Scatter Plot**

Select file `ftcanmax.R`

Select **line** option

x-axis variable: `Year`

y-axis variable: `Prec`

Click on **OK**

#### (iii) Fit GEV distribution to Fort Collins data

##### **Analyze/Generalized Extreme Value (GEV) Distribution**

Data Object: `ftcanmax.R`

Optimization Method: **Nelder-Mead**

Response: `Prec`

Plot diagnostics: **check**

Click on **OK**

Note: Fitted output saved in `gev.fit1`

(iv) *Estimate return level for Fort Collins data*

**Analyze/Parameter confidence interval/GEV fit**

Data Object: `ftcanmax.R`

Select a fit: `gev.fit1`

m-year return level: **100**

Return level: **check**

Shape parameter: **uncheck**

Lower limit: **300**

Upper limit: **900**

Plot profile likelihoods: **yes**

Click on **OK**

(3.2) *Generalized Pareto (GP)/Distribution tails**Phoenix daily minimum temperature*(i) *Read Phoenix data into extRemes***File/Read Data**

Select file: Tphap.R

File Type: **R source**

Save as (in R): Tphap.R

Click on **OK**Note: Creates extRemes data object Tphap.R(ii) *Negate minimum temperature***File/Transform Data/Negative**

Data Object: Tphap.R

Variables to Transform: MinT

Click on **OK**Note: Creates new variable MinT.neg(iii) *Fit GP distribution to negated data***Analyze/Generalized Pareto Distribution (GPD)**

Data Object: Tphap.R

Response: MinT.neg

Plot diagnostics: **check**Method: **Nelder-Mead**Threshold: **-73**Number of obs per year: **62**Click on **OK**Note: Fitted output saved in gpd.fit1

(iv) *Identify threshold for negated data*

**Plot/Fit Threshold ranges (GPD)**

Data Object: Tphap.R

Variable: MinT.neg

Threshold Ranges (GPD):

Minimum Threshold: **-78**

Maximum Threshold: **-68**

Number of Thresholds: **11**

Click on **OK**

(v) *Decluster negated minimum temperature*

**File/Decluster**

Data Object: Tphap.R

Variable to Decluster: MinT.neg

Decluster by: Year

Threshold(s): **-73**

Run Length: **1**

Click on **OK**

Note: Creates new variable MinT.neg.u-73r1dcbyYear

(vi) *Fit GP distribution to declustered data*

**Analyze/Generalized Pareto Distribution (GPD)**

Data Object: Tphap.R

Response: MinT.neg.u-73r1dcbyYear

Plot diagnostics: **check**

Method: **Nelder-Mead**

Threshold: **-73**

Number of obs per year: **62**

Click on **OK**

Note: Fitted output saved in gpd.fit2

(4) *Exercise*

*Economic damage from hurricanes (fitting GP distribution to upper tail)*

- (i) Read data file `damage.R` into `extRemes`
- (ii) Fit GP distribution to upper tail of response variable (economic damage from hurricanes):  
    Dam

To obtain more information about this data set, use R command:

```
help(damage)
```

Hints:

- (i) Need to identify appropriate threshold for damage data
- (ii) Should be no need to decluster damage data
- (iii) Number of observations per year ( $\approx 2$  damaging hurricanes per year)

## PART 2 OF TUTORIAL ON EXTREMES TOOLKIT

### (1) *Using Covariates in Extremes Toolkit*

*Simulation from GEV distribution with trend*

#### (i) *Simulate data*

#### **File/Simulate Data/Generalized Extreme Value (GEV)**

GEV parameters:

Location parameter ( $\mu$ ): **0**

Scale parameter ( $\sigma$ ): **1**

Shape parameter ( $\xi$ ): **0.2**

Trend (in location parameter): **0.01**

Sample size: **100**

Save as: `gevtrend.R`

Click on **Generate**

Note: Creates `extRemes` data object `gevtrend.R`

#### (ii) *Fit GEV with no trend to simulated data*

#### **Analyze/Generalized Extreme Value (GEV) Distribution**

Data Object: `gevtrend.R`

Optimization Method: **Nelder-Mead**

Response: `gev.sim`

Plot diagnostics: **check**

Click on **OK**

Note: Fitted output saved in `gev.fit1`

(iii) *Fit GEV with linear trend in location parameter to simulated data*

**Analyze/Generalized Extreme Value (GEV) Distribution**

Data Object: `gevtrend.R`

Optimization Method: **Nelder-Mead**

Response: `gev.sim`

Location parameter ( $\mu$ ): **obs**

Link: **identity**

Plot diagnostics: **check**

Click on **OK**

Note: Fitted output saved in `gev.fit2`

(iv) *Perform likelihood-ratio test to test for trend in location parameter*

**Analyze/Likelihood-ratio test**

Data Object: `gevtrend.R`

Significance level ( $\alpha$ ): **0.05**

Select base fit (M0): `gev.fit1`

Select comparison fit (M1): `gev.fit2`

Click on **OK**

## (2) *Climate Change Applications*

*Block maxima approach: Phoenix summer minimum temperature (trend)*

To obtain more information about this data set, use R command: `help(HEAT)`

### (i) *Read Phoenix data into extRemes*

#### **File/Read Data**

Select file `HEAT.R`

File Type: **R source**

Save as (in R): `HEAT.R`

Click on **OK**

Note: Creates `extRemes` data object `HEAT.R`

### (ii) *Check data*

#### **Plot/Scatter Plot**

Select file `HEAT.R`

Select **line** option

x-axis variable: `Year`

y-axis variable: `Tmin`

Click on **OK**

### (iii) *Negate minimum temperature*

#### **File/Transform Data/Negative**

Data Object: `HEAT.R`

Variables to Transform: `Tmin`

Click on **OK**

Note: Creates new variable `Tmin.neg`

(iv) *Transform time variable*

**File/Transform Data/Affine Transformation**

Data Object: HEAT.R

Variables to Transform: Year

c (subtract): **47**

b (divide by): **1**

Click on **OK**

Note: Creates new variable Year.c47b1

(v) *Fit GEV with no trend to negated data*

**Analyze/Generalized Extreme Value (GEV) Distribution**

Data Object: HEAT.R

Optimization Method: **Nelder-Mead**

Response: Tmin.neg

Plot diagnostics: **check**

Click on **OK**

Note: Fitted output saved in gev.fit1

(vi) *Fit GEV with linear trend in location parameter to negated data*

**Analyze/Generalized Extreme Value (GEV) Distribution**

Data Object: HEAT.R

Optimization Method: **Nelder-Mead**

Response: Tmin.neg

Location parameter (mu): Year.c47b1

Link: **identity**

Plot diagnostics: **check**

Click on **OK**

Note: Fitted output saved in gev.fit2

(vii) *Perform likelihood-ratio test to test for trend in location parameter*

**Analyze/Likelihood-ratio test**

Data Object: HEAT.R

Significance level (alpha): **0.05**

Select base fit (M0): `gev.fit1`

Select comparison fit (M1): `gev.fit2`

Click on **OK**

(viii) *Fit GEV with linear trends in both location parameter and log(scale parameter) to negated data*

**Analyze/Generalized Extreme Value (GEV) Distribution**

Data Object: HEAT.R

Optimization Method: **Nelder-Mead**

Response: `Tmin.neg`

Location parameter (mu): `Year.c47b1`

Link: **identity**

Scale parameter (sigma): `Year.c47b1`

Link: **log**

Plot diagnostics: **check**

Click on **OK**

Note: Fitted output saved in `gev.fit3`

(ix) *Perform likelihood-ratio test for trend in log(scale parameter)*

**Analyze/Likelihood-ratio test**

Data Object: HEAT.R

Significance level (alpha): **0.05**

Select base fit (M0): `gev.fit2`

Select comparison fit (M1): `gev.fit3`

Click on **OK**

(3) *Exercise*

*Port Jervis winter maximum temperature (fitting GEV distribution with Arctic Oscillation as covariate)*

(i) Read data file `PORTw.R` into `extRemes`

Response variable (winter maximum temperature): `TMX1`

Covariate/predictor variable (Arctic Oscillation index): `AOindex`

To obtain more information about this data set, use R command: `help(PORTw)`

(ii) Fit GEV distribution to winter maximum temperature with

(a) No covariates

(b) Location parameter as linear function of Arctic Oscillation

(c) Location parameter and  $\log(\text{scale parameter})$  as linear functions of Arctic Oscillation

(iii) Perform likelihood-ratio tests for effect of Arctic Oscillation on location and  $\log(\text{scale parameter})$  of GEV distribution for maximum temperature

(iv) For best model (based on outcome of likelihood-ratio tests), interpret effect of Arctic Oscillation in terms of return levels of GEV distribution for maximum temperature

Hint: An expression for the return level with return period  $1/p$ , denoted by  $x(p)$ , of a GEV distribution is given by

$$x(p) = \mu - (\sigma/\zeta) \{1 - [-\log(1 - p)]^{-\zeta}\}, \quad \zeta \neq 0$$

where  $\mu$ ,  $\sigma$ , and  $\zeta$  denote the location, scale, and shape parameters, respectively.