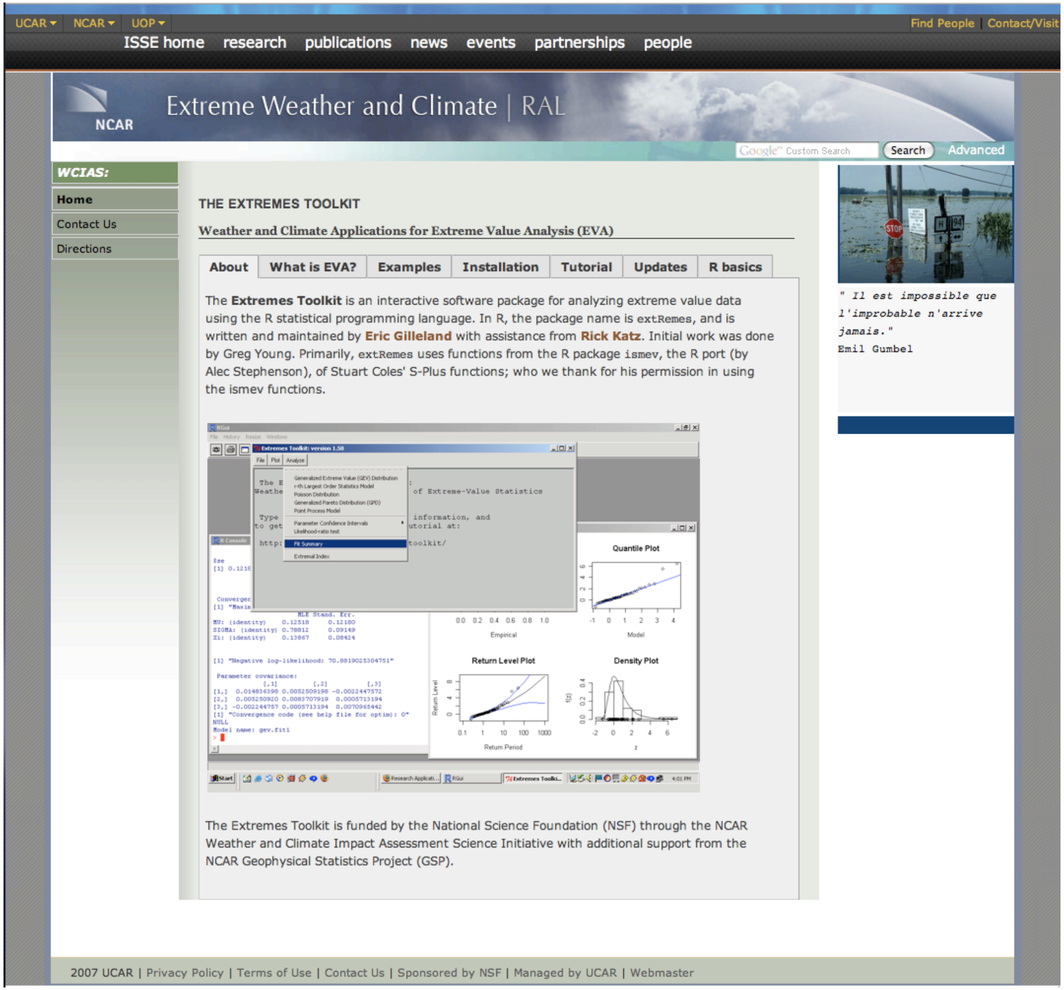
Extreme Value Theory, Extreme Temperatures, and demonstration of `extRemes`

Eric Gilleland,
National Center for
Atmospheric Research,
Boulder, Colorado, U.S.A.
http://www.ral.ucar.edu/staff/ericg

Third Biannual NCAR Workshop on Climate and Health
Effects of Heat Waves and Air Pollution
15 July 2009

Extremes Toolkit (extRemes) Web Page

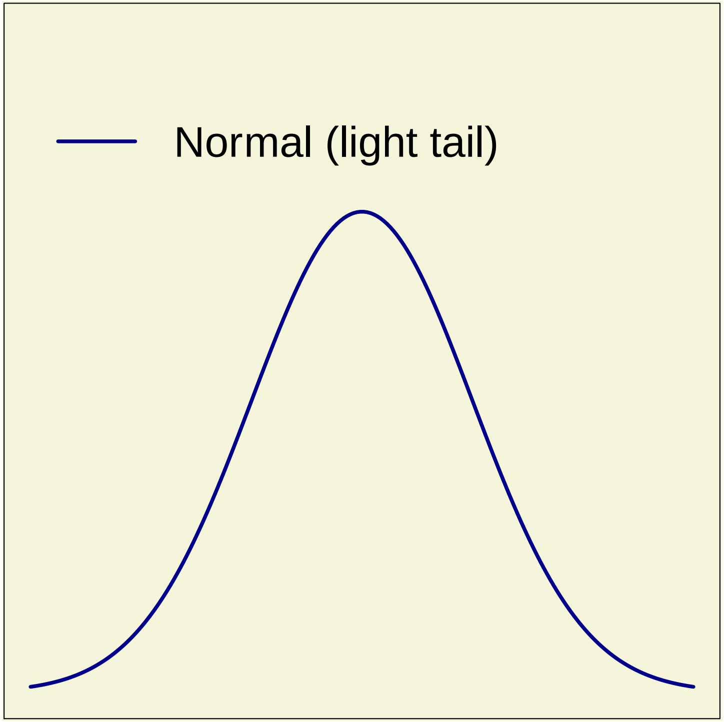
http://www.isse.ucar.edu/extremevalues/evtk.html



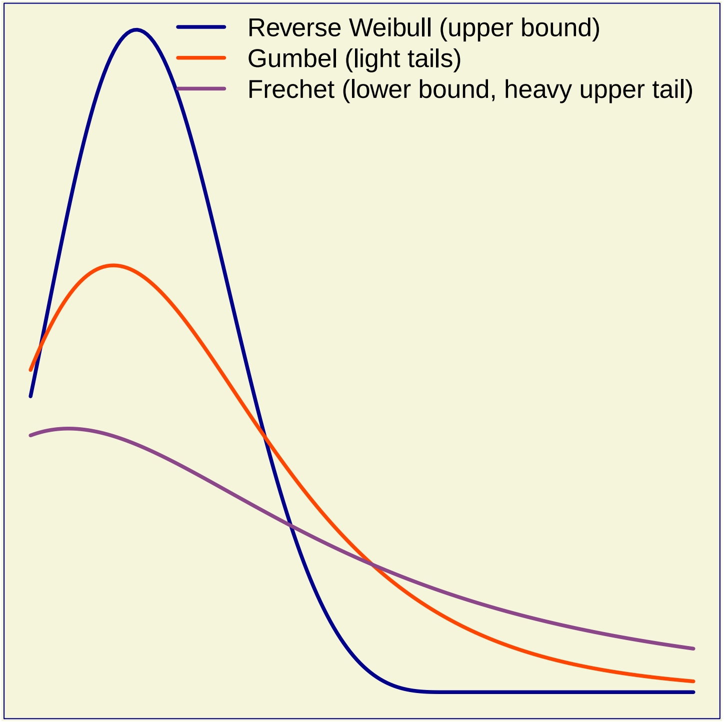
Background on Extreme Value Analysis (EVA)

Motivation

Central Limit Theorem



Extremal Types Theorem



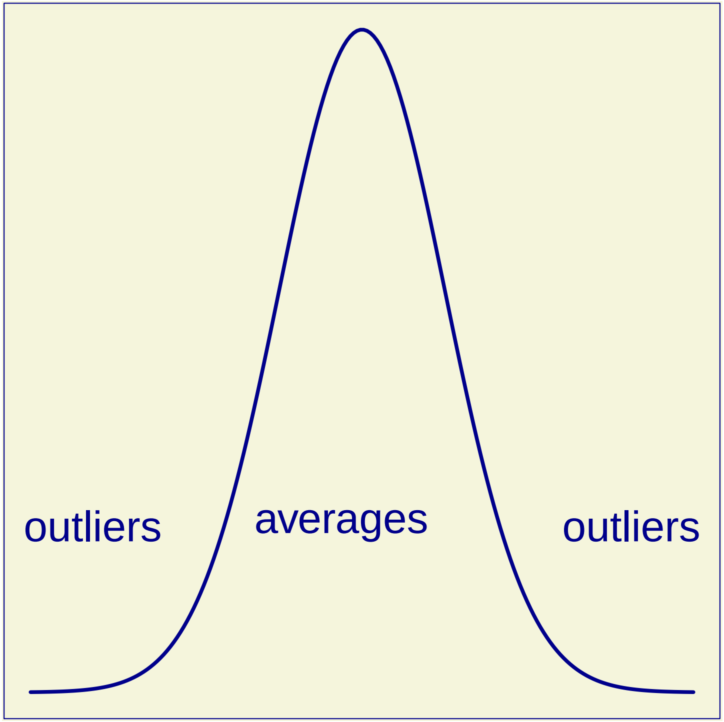
Sums/Averages

Maxima/Minima (threshold excesses)

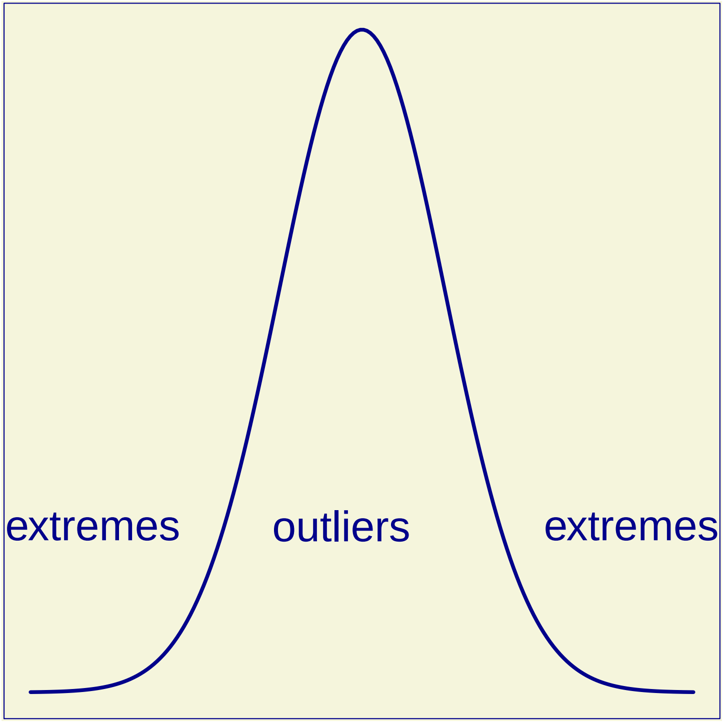
Background on Extreme Value Analysis (EVA)

Motivation

Sums/Averages

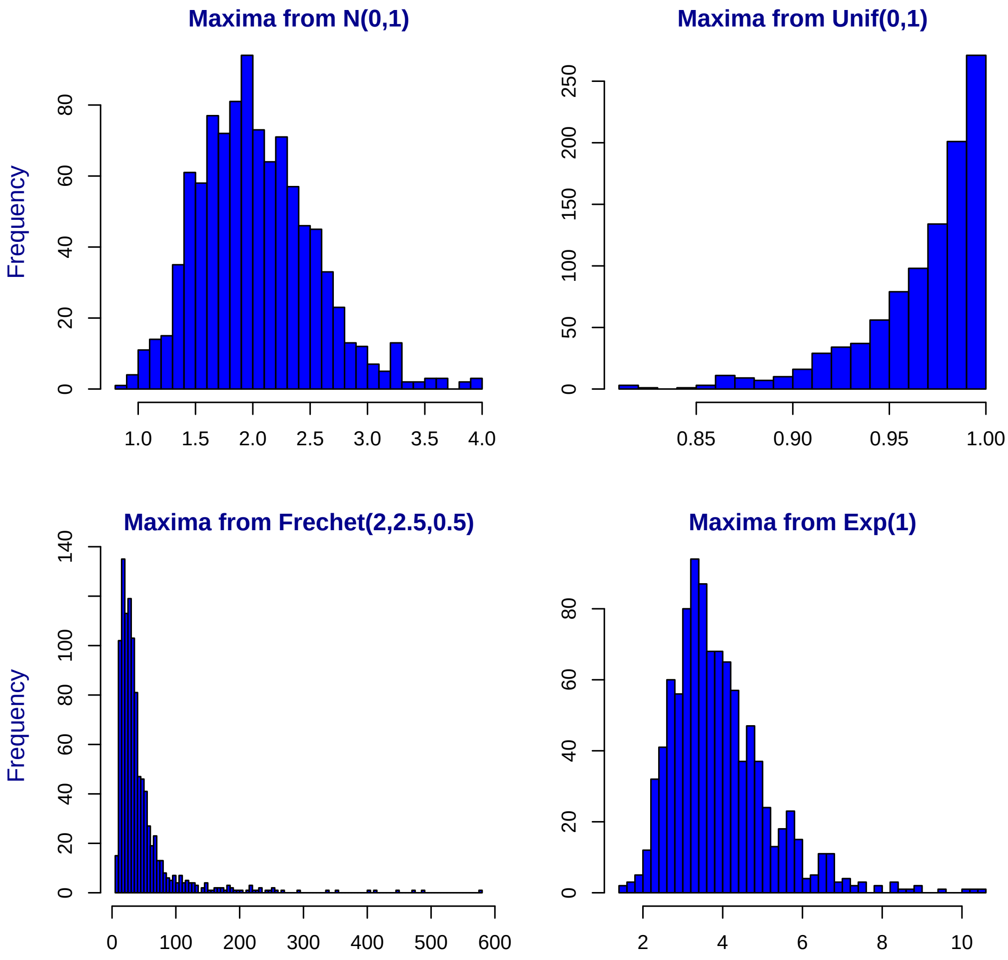


Maxima (Minima)/Threshold Excesses



Background on Extreme Value Analysis (EVA)

Simulations



Background on Extreme Value Analysis (EVA)

Extremal Types Theorem

Let $X_1, \ldots, X_n$ be a sequence of independent and identically distributed (iid) random variables with common distribution function, F. Want to know the distribution of

$$M_n = \max\{X_1, \ldots, X_n\}.$$

Example: $X_1, \ldots, X_n$ could represent hourly precipitation, daily ozone concentrations, daily average temperature, etc. Interest for now is in maxima of these variables over particular blocks of time.

Background on Extreme Value Analysis (EVA)

Extremal Types Theorem

If interest is in the minimum over blocks of data
(e.g., monthly minimum temperature), then note that

$$\min\{X_1, \ldots, X_n\} = -\max\{-X_1, \ldots, -X_n\}$$

Therefore, we can focus on the maxima.

Background on Extreme Value Analysis (EVA)

Extremal Types Theorem

Could try to derive the distribution for M_n exactly for all n as follows.

$$\begin{aligned}
\Pr\{M_n \leq z\} &= \Pr\{X_1 \leq z, \ldots, X_n \leq z\} \\
&\stackrel{\text{indep.}}{=} \Pr\{X_1 \leq z\} \times \cdots \times \Pr\{X_n \leq z\} \\
&\stackrel{\text{ident. dist.}}{=} \{F(z)\}^n.
\end{aligned}$$

Background on Extreme Value Analysis (EVA)

Extremal Types Theorem

Could try to derive the distribution for M_n exactly for all n as follows.

$$
\begin{aligned}
\Pr\{M_n \leq z\} &= \Pr\{X_1 \leq z, \ldots, X_n \leq z\} \\
&\overset{\text{indep.}}{=} \Pr\{X_1 \leq z\} \times \cdots \times \Pr\{X_n \leq z\} \\
&\overset{\text{ident. dist.}}{=} \{F(z)\}^n.
\end{aligned}
$$

But! If F is not known, this is not very helpful because small discrepancies in the estimate of F can lead to large discrepancies for F^n.

Background on Extreme Value Analysis (EVA)

Extremal Types Theorem

Could try to derive the distribution for M_n exactly for all n as follows.

$$\begin{aligned}
\Pr\{M_n \le z\} &= \Pr\{X_1 \le z, \ldots, X_n \le z\} \\
&\stackrel{\text{indep.}}{=} \Pr\{X_1 \le z\} \times \cdots \times \Pr\{X_n \le z\} \\
&\stackrel{\text{ident. dist.}}{=} \{F(z)\}^n.
\end{aligned}$$

But! If F is not known, this is not very helpful because small discrepancies in the estimate of F can lead to large discrepancies for F^n.

Need another strategy!

Background on Extreme Value Analysis (EVA)

Extremal Types Theorem

If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$\Pr\left\{\frac{M_n - b_n}{a_n} \leq z\right\} \longrightarrow G(z) \text{ as } n \longrightarrow \infty,$$

where G is a non-degenerate distribution function, then G belongs to one of the following three types.

Background on Extreme Value Analysis (EVA)

Extremal Types Theorem

I. Gumbel

$$G(z) = \exp\left\{-\exp\left[-\left(\frac{z-b}{a}\right)\right]\right\}, \quad -\infty < z < \infty$$

II. Fréchet

$$G(z) = \begin{cases} 0, & z \leq b, \\ \exp\left\{-\left(\frac{z-b}{a}\right)^{-\alpha}\right\}, & z > b; \end{cases}$$

III. Weibull

$$G(z) = \begin{cases} \exp\left\{-\left[-\left(\frac{z-b}{a}\right)^{\alpha}\right]\right\}, & z < b, \\ 1, & z \geq b \end{cases}$$

with parameters a, b and $\alpha > 0$.

Background on Extreme Value Analysis (EVA)

Extremal Types Theorem

The three types can be written as a single family of distributions, known as the generalized extreme value (GEV) distribution.

$$G(z) = \exp\left\{ -\left[1 + \xi \left(\frac{z-\mu}{\sigma} \right) \right]_{+}^{-1/\xi} \right\},$$

where $y_+ = \max\{y, 0\}$, $-\infty < \mu, \xi < \infty$ and $\sigma > 0$.

Background on Extreme Value Analysis (EVA)

GEV distribution

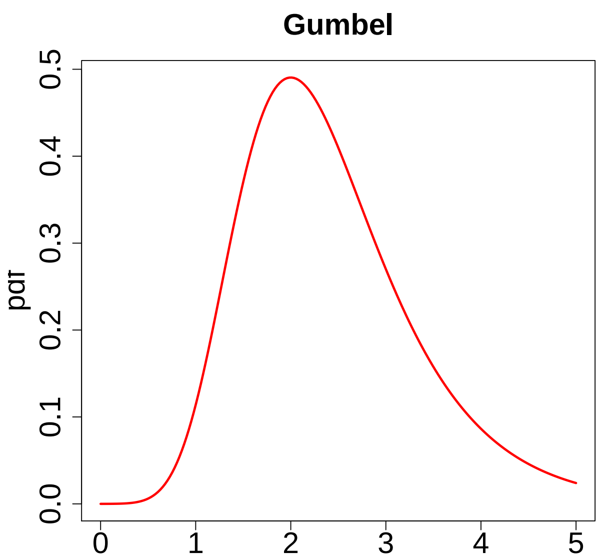
Three parameters: location (μ), scale (σ) and shape (ξ).

1. $\xi = 0$ (Gumbel type, limit as $\xi \longrightarrow 0$)
2. $\xi > 0$ (Fréchet type)
3. $\xi < 0$ (Weibull type)

Background on Extreme Value Analysis (EVA)

Gumbel type

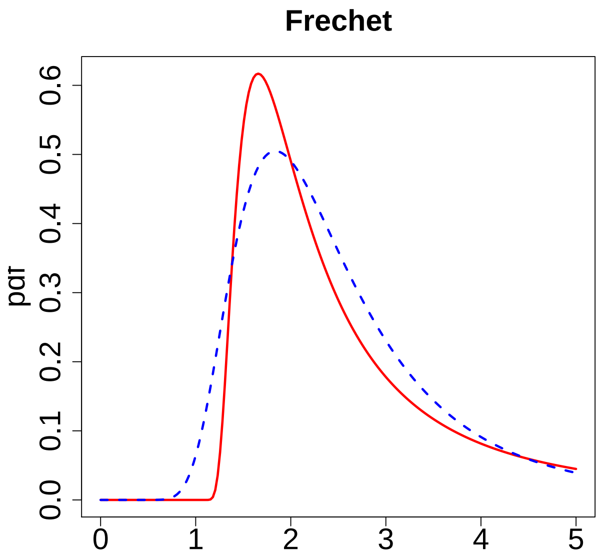
- Light tail
- Domain of attraction for many common distributions (e.g., normal, lognormal, exponential, gamma)



Background on Extreme Value Analysis (EVA)

Fréchet type

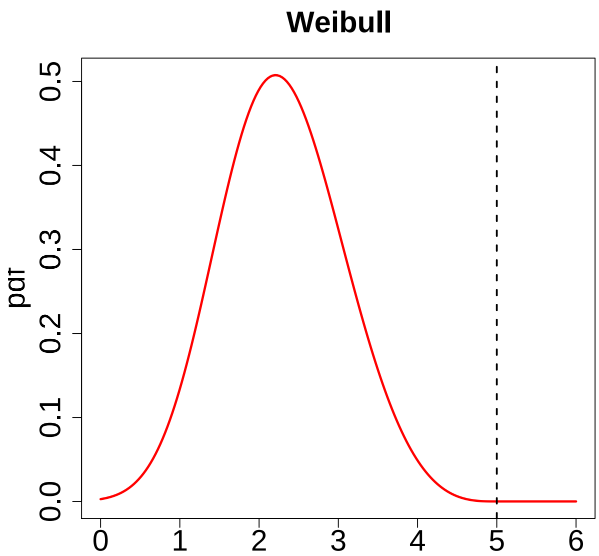
- Heavy tail
- $\mathcal{E}[X^r] = \infty$ for $r \geq 1/\xi$ (i.e., infinite variance if $\xi \geq 1/2$)
- Of interest for precipitation, streamflow, economic impacts



Background on Extreme Value Analysis (EVA)

Weibull type

- Bounded upper tail at $\mu - \frac{\sigma}{\xi}$
- Of interest for temperature, wind speed, sea level



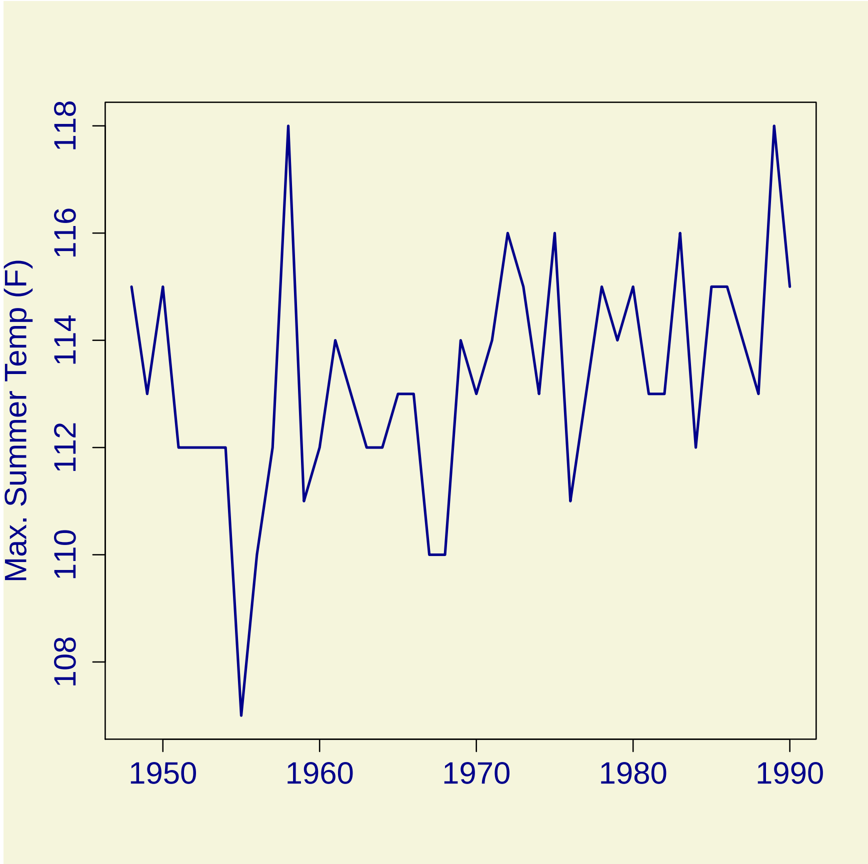
Background on Extreme Value Analysis (EVA)

Normal vs. GEV

$\Pr\{X > \cdot\}$	1	2	4	8	16	32
N(0,1)	0.16	0.02	$< 10^{-4}$	$< 10^{-15}$	$< 10^{-50}$	$< 10^{-200}$
Gumbel(0,1)	0.31	0.13	0.02	$< 10^{-3}$	$< 10^{-6}$	$< 10^{-13}$
Fréchet(0,1,0.5)	0.36	0.22	0.11	0.04	0.01	0.003
Weibull(0,1,-0.5)	0.22	0	0	0	0	0

Example

Phoenix Sky Harbor airport summer (July–August) 1948–1990 maximum (and minimum) temperature (°F)



Source: U.S. National Weather Service Forecast office at the Phoenix Sky Harbor Airport (via `extRemes`).

Phoenix Sky Harbor airport summer (July–August) 1948–1990 maximum (and minimum) temperature (°F)

Demo: Reading in the `HEAT` data to `extRemes`

Phoenix Sky Harbor airport summer (July–August) 1948–1990 maximum (and minimum) temperature (°F)

Demo:

`ls, class, names, colnames, dim, ...`

Phoenix Sky Harbor airport summer (July–August) 1948–1990 maximum (and minimum) temperature (°F)

Demo: Scatter (line) plot using `extRemes`

Phoenix Sky Harbor airport summer (July–August) 1948–1990 maximum (and minimum) temperature (°F)

Demo: take the negative of the minimum temperatures.

Example

Fort Collins, Colorado precipitation

What sort of extreme temperatures can we expect in Phoenix?

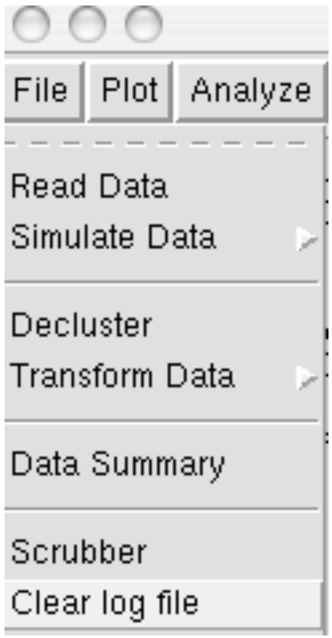
- Assume no long-term trend emerges (for now).
- Using annual maxima removes effects of annual trend in analysis.
- Annual Maxima/(negative) Minima fit to GEV.

Demo: Fitting a (stationary) GEV to maxima and (negative) minima.

Command-line Code Executed

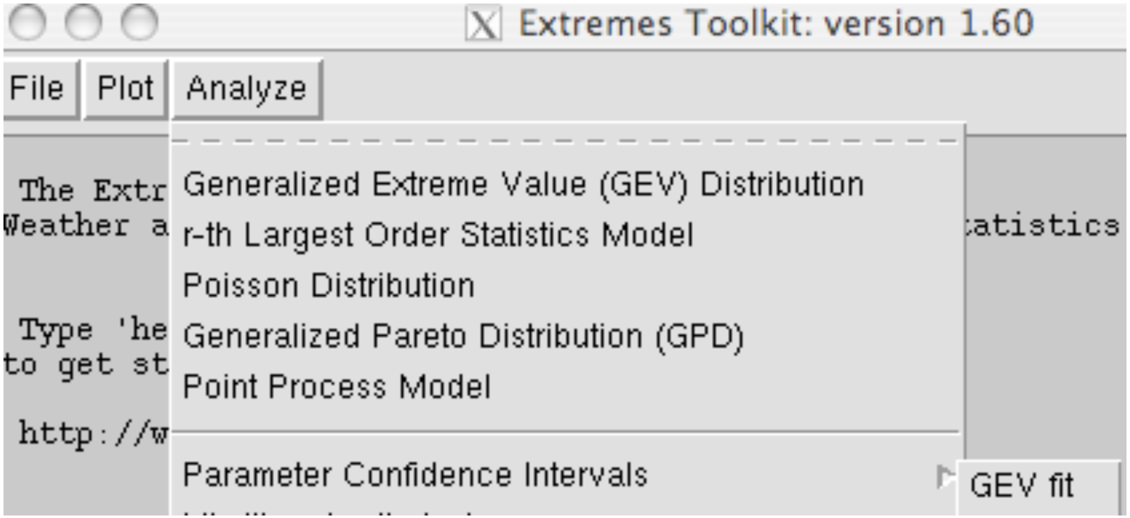
To see the (underlying) code used to execute this fit, look at the `extRemes.log` file found in your working R directory (use `getwd()` to find this directory).

Should periodically clear this file because it will get larger as more commands are executed.



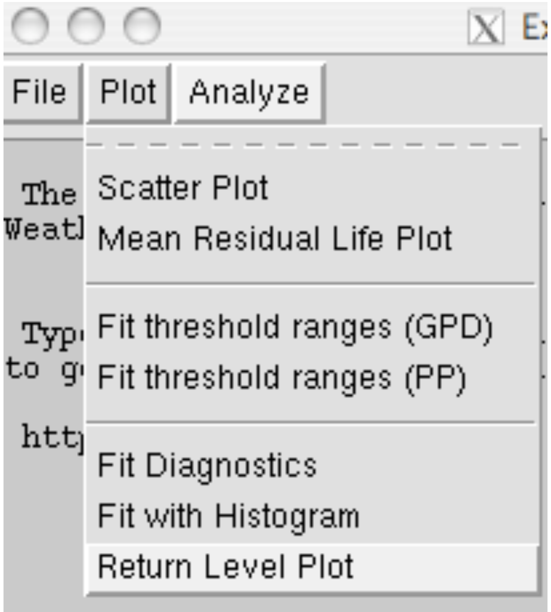
Phoenix Sky Harbor airport summer (July–August) 1948–1990 maximum (and minimum) temperature (°F)

Demo: Estimate 95% CI's for shape parameter using profile likelihood.



Phoenix Sky Harbor airport summer (July–August) 1948–1990 maximum (and minimum) temperature (°F)

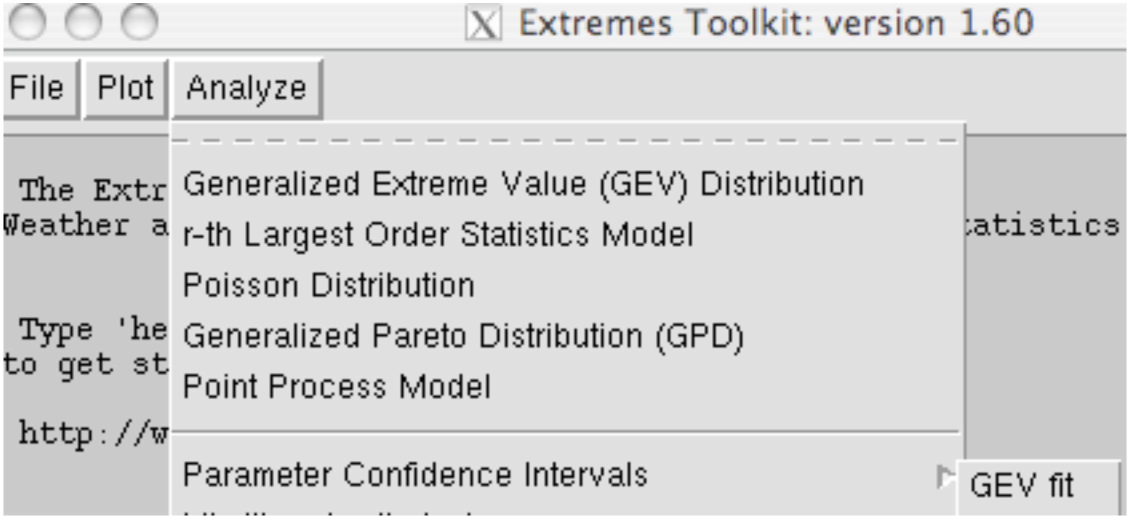
Demo: Return Levels



Phoenix Sky Harbor airport summer (July–August) 1948–1990 maximum (and minimum) temperature (°F)

Return Levels

Demo: profile likelihood to determine CI's for longer return periods.



Peaks Over Thresholds (POT) Approach

Let $X_1, X_2, \ldots$ be an iid sequence of random variables, again with marginal distribution, F. Interest is now in the conditional probability of X's exceeding a certain value, given that X already exceeds a sufficiently large threshold, u.

$$\Pr\{X > u + y | X > u\} = \frac{1 - F(u+y)}{1 - F(u)}, \; y > 0$$

Once again, if we know F, then the above probability can be computed. Generally not the case in practice, so we turn to a broadly applicable approximation.

Peaks Over Thresholds (POT) Approach

If $\Pr\{\max\{X_1, \ldots, X_n\} \leq z\} \approx G(z)$, where

$$G(z) = \exp\left\{ -\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi} \right\}$$

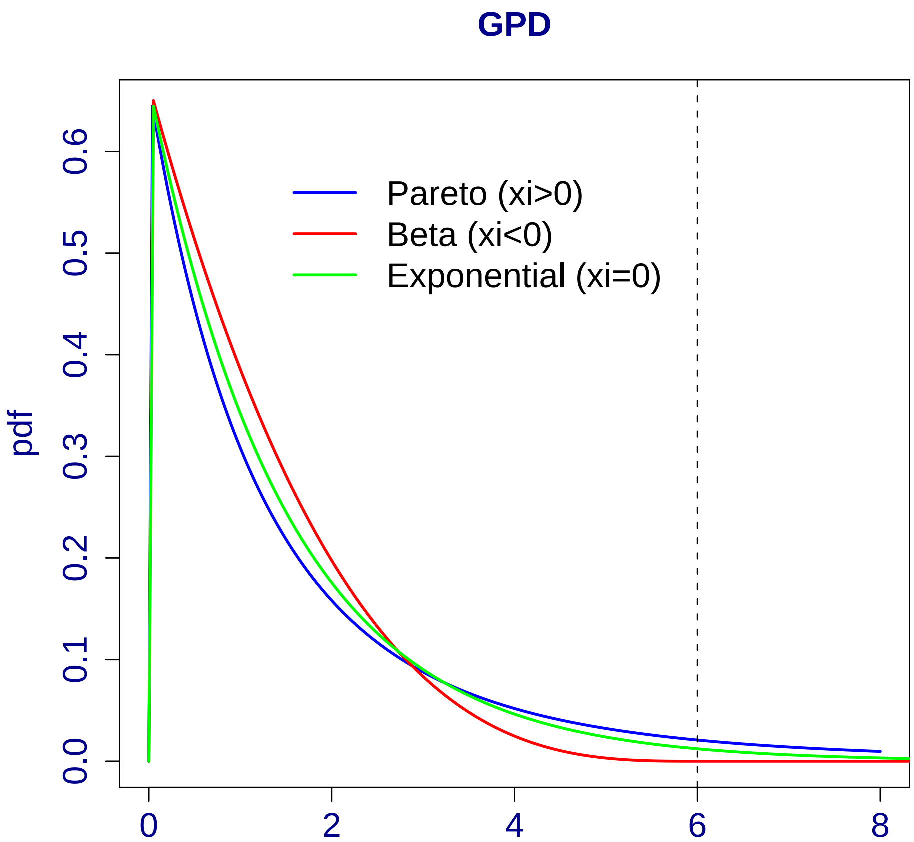
for some μ, ξ and $\sigma > 0$, then for sufficiently large u, the distribution $[X - u | X > u]$, is approximately the generalized Pareto distribution (GPD). Namely,

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)_+^{-1/\xi}, \; y > 0,$$

with $\tilde{\sigma} = \sigma + \xi(u - \mu)$ (σ, ξ and μ as in $G(z)$ above).

Peaks Over Thresholds (POT) Approach

- Pareto type ($\xi > 0$)
  heavy tail
- Beta type ($\xi < 0$)
  bounded above at
  $u - \sigma/\xi$
- Exponential type ($\xi = 0$)
  light tail

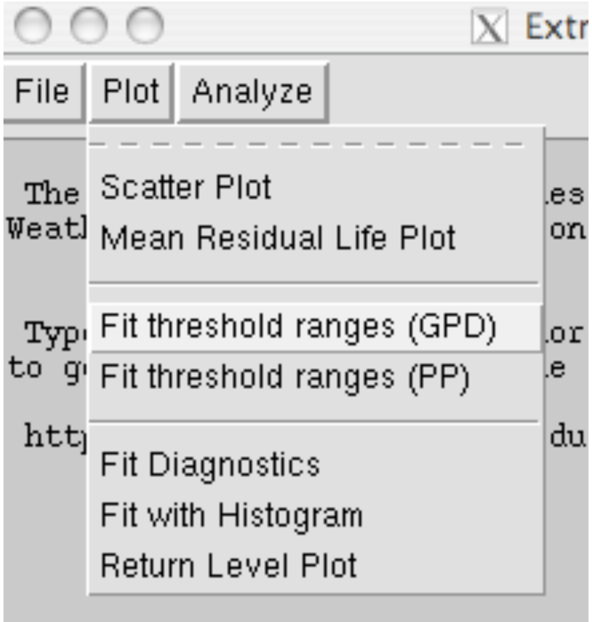


Peaks Over Thresholds (POT) Approach

Choosing a threshold

Variance/bias trade-off

Low threshold allows for more data (low variance).
Theoretical justification for GPD requires a high threshold (low bias).



Choosing a threshold

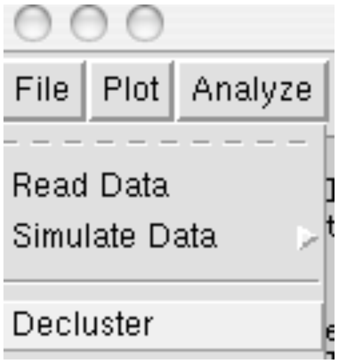
Demo: Choosing a threshold.

Peaks Over Thresholds (POT) Approach

Dependence above threshold

Often, threshold excesses are *not* independent. For example, a hot day is likely to be followed by another hot day.

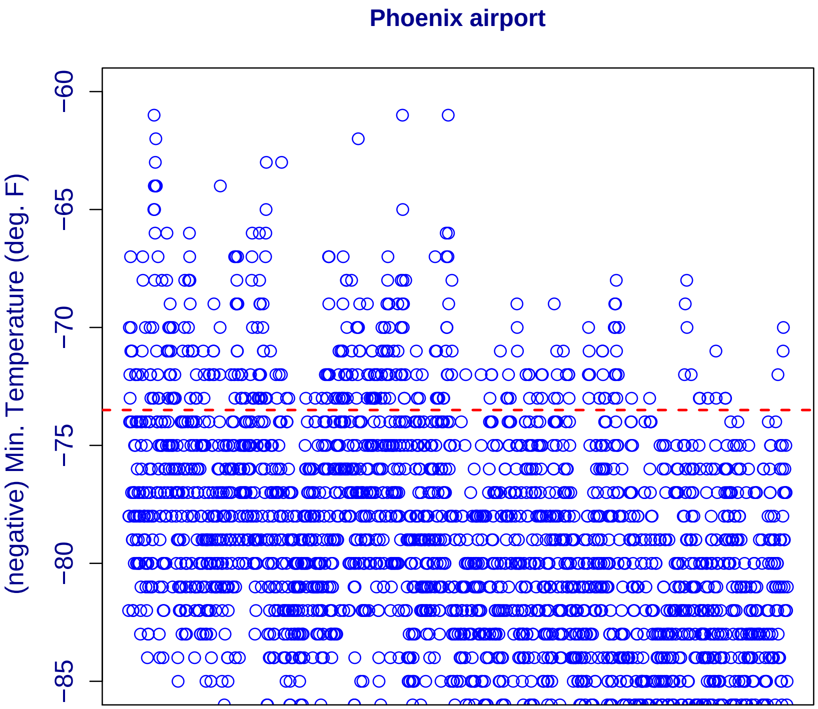
Various procedures to handle dependence.



- Model the dependence.
- De-clustering (e.g., runs de-clustering).
- Resampling to estimate standard errors (avoid tossing out information about extremes).

Peaks Over Thresholds (POT) Approach

Dependence above threshold



Phoenix (airport) minimum temperature (°F).

July and August 1948–1990.

Urban heat island (warming trend as cities grow).

Model lower tail as upper tail after negation.

Peaks Over Thresholds (POT) Approach

Dependence above threshold

Fit without de-clustering.
$\hat{\sigma} \approx 3.93$
$\hat{\xi} \approx -0.25$

With runs de-clustering (r=1).
$\hat{\sigma} \approx 4.21$
$\hat{\xi} \approx -0.25$

Peaks Over Thresholds (POT) Approach

Point Process: *frequency and intensity of threshold excesses*

Event is a threshold excess (i.e., $X > u$).

Frequency of occurrence of an event (rate parameter), $\lambda > 0$.

$\Pr\{\text{no events in } [0, T]\} = e^{-\lambda T}$

Mean number of events in $[0, T] = \lambda T$.

GPD for excess over threshold (intensity).

Peaks Over Thresholds (POT) Approach

Point Process: *frequency and intensity of threshold excesses*

Relation of parameters of GEV(μ,σ,ξ) to parameters of point process (λ,σ^*,ξ).

- Shape parameter, ξ, identical.
- $\log \lambda = -\frac{1}{\xi} \log \left(1 + \xi \frac{u-\mu}{\sigma}\right)$
- $\sigma^* = \sigma + \xi(u - \mu)$

More detail: Time scaling constant, h. For example, for annual maximum of daily data, $h \approx 1/365.25$. Change of time scale, h, for GEV(μ,σ,ξ) to h'

$$\sigma' = \sigma \left(\frac{h}{h'}\right)^{\xi} \text{ and } \mu' = \mu + \frac{1}{\xi}\left\{\sigma'\left[1 - \left(\frac{h}{h'}\right)^{-\xi}\right]\right\}$$

Peaks Over Thresholds (POT) Approach

Point Process: *frequency and intensity of threshold excesses*

Two ways to estimate PP parameters

- Orthogonal approach (estimate frequency and intensity separately).

  Convenient to estimate.

  Difficult to interpret in presence of covariates.

- GEV re-parameterization (estimate both simultaneously).

  More difficult to estimate.

  Interpretable even with covariates.

Peaks Over Thresholds (POT) Approach

Point Process: *frequency and intensity of threshold excesses*

Daily (negative) minimum temperature (°F) July–August 1948–1990 at Phoenix Sky Harbor Airport (`Tphap`

Analyze daily data instead of just annual maxima (ignoring annual cycle for now).

Orthogonal Approach

$$\hat{\lambda} = 62 \text{ days per year} \cdot \frac{\text{No. } X_i > -73}{\text{No. } X_i} \approx 6.1 \text{ per year}$$

$$\hat{\sigma}^* \approx 3.93,\ \hat{\xi} \approx -0.25$$

Demo: Estimate using GUI windows (Transform to indicator above threshold, then fit Poisson).

Peaks Over Thresholds (POT) Approach

Point Process: *frequency and intensity of threshold excesses*

Fort Collins, Colorado daily precipitation

Analyze daily data instead of just annual maxima
(ignoring annual cycle for now).

Point Process

$$\hat{\mu} \approx -63.68 \text{ (i.e., } 63.68)$$
$$\hat{\sigma} = 1.62$$
$$\hat{\xi} \approx -0.25$$

Risk Communication Under Stationarity

Unchanging climate

Return level, z_p, is the value associated with the **return period**, $1/p$. That is, z_p is the level expected to be exceeded on average once every $1/p$ years.

That is, Return level, z_p, with $1/p$-year return period is

$$z_p = F^{-1}(1-p).$$

For example, $p = 0.01$ corresponds to the 100-year return period.

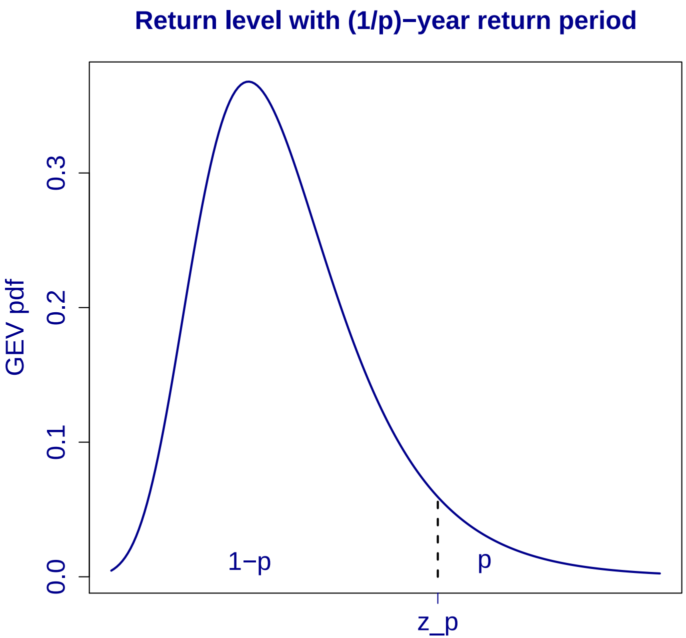
Easy to obtain from GEV and GP distributions (stationary case).

Risk Communication Under Stationarity

Unchanging climate

For example, GEV return level is given by

$$z_p = \mu - \frac{\sigma}{\xi}\left[1 - (-\log(1-p))\right]^{-\xi}$$



Similar for GPD, but must take λ into account.

Non-Stationarity

Sources

- Trends:
  climate change: trends in frequency and intensity of extreme weather events.
- Cycles:
  Annual and/or diurnal cycles often present in meteorological variables.
- Other.

Non-Stationarity

Theory

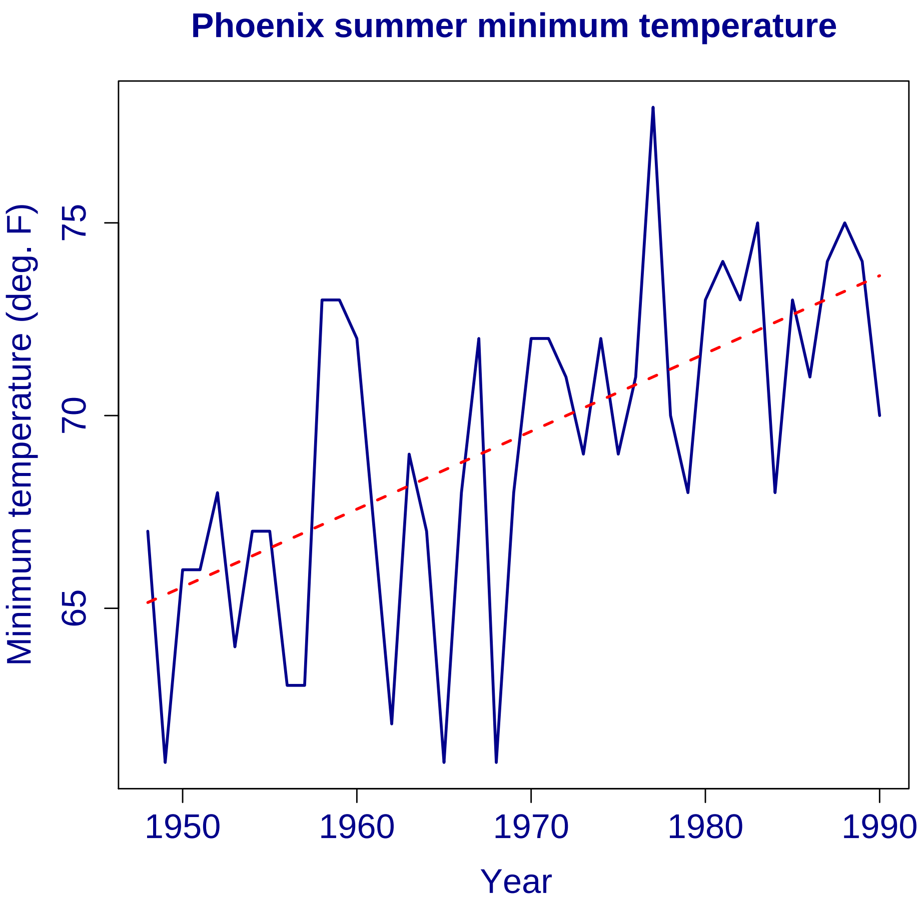
No general theory for non-stationary case.

Only limited results under restrictive conditions.

Can introduce covariates in the distribution parameters.

Non-Stationarity

Phoenix minimum temperature



Phoenix minimum temperature

Recall: $\min\{X_1, \ldots, X_n\} = -\max\{-X_1, \ldots, -X_n\}$.

Assume summer minimum temperature in year $t = 1, 2, \ldots$ has GEV distribution with:

$$\mu(t) = \mu_0 + \mu_1 \cdot t$$

$$\log \sigma(t) = \sigma_0 + \sigma_1 \cdot t$$

$$\xi(t) = \xi$$

Non-Stationarity

Phoenix minimum temperature

Note: To convert back to $\min\{X_1, \ldots, X_n\}$,
change sign of location parameters. But note that model is
$\Pr\{-X \le x\} = \Pr\{X \ge -x\} = 1 - F(-x)$.

$$\hat{\mu}(t) \approx 66.170 + 0.196t$$

$$\log \hat{\sigma}(t) \approx 1.338 - 0.009t$$

$$\hat{\xi} \approx -0.21$$

Likelihood ratio test

for $\mu_1 = 0$ (p-value $< 10^{-5}$),

for $\sigma_1 = 0$ (p-value ≈ 0.366).

Non-Stationarity

Phoenix minimum temperature

Model Checking. Found the best model from a range of models, but is it a good representation of the data? Transform data to a common distribution, and check the qq-plot.

1. Non-stationary GEV to exponential

$$\varepsilon_t = \left\{ 1 + \frac{\hat{\xi}(t)}{\hat{\sigma}(t)} [X_t - \hat{\mu}(t)] \right\}^{-1/\hat{\xi}(t)}$$

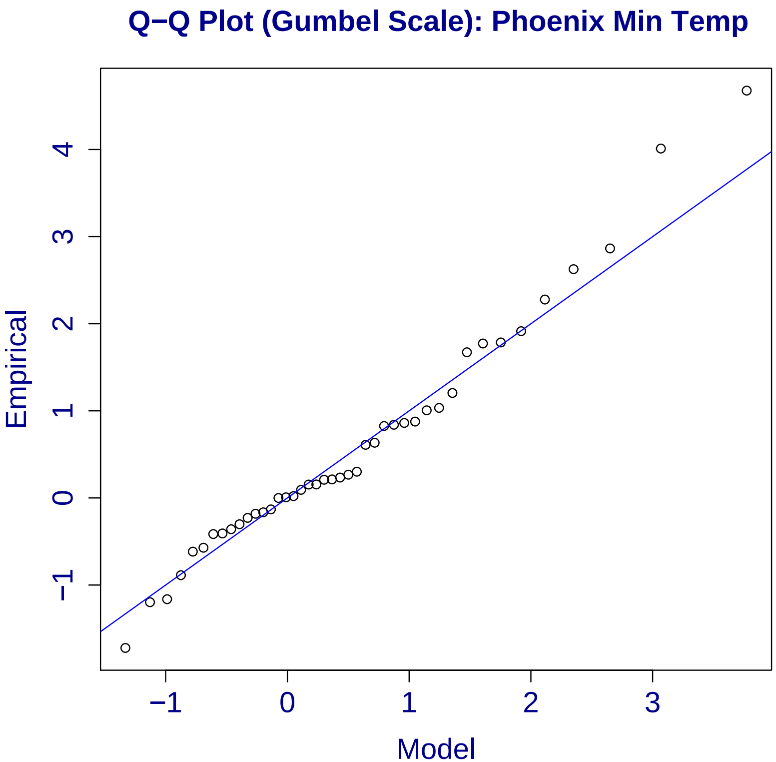
2. Non-stationary GEV to Gumbel (used by `ismev`/`extRemes`)

$$\varepsilon_t = \frac{1}{\hat{\xi}(t)} \log \left\{ 1 + \hat{\xi}(t) \left(\frac{X_t - \hat{\mu}(t)}{\hat{\sigma}(t)} \right) \right\}$$

Non-Stationarity

Phoenix minimum temperature

Model Checking. Found the best model from a range of models, but is it a good representation of the data? Transform data to a common distribution, and check the qq-plot.



Non-Stationarity

Physically based covariates

Winter maximum daily temperature at Port Jervis, New York

Let $X_1, \ldots, X_n$ be the winter maximum temperatures, and $Z_1, \ldots, Z_n$ the associated Arctic Oscillation (AO) winter index. Given $Z = z$, assume conditional distribution of winter maximum temperature is GEV with parameters

$$\mu(z) = \mu_0 + \mu_1 \cdot z$$

$$\log \sigma(z) = \sigma_0 + \sigma_1 \cdot z$$

$$\xi(z) = \xi$$

Non-Stationarity

Physically based covariates

Winter maximum daily temperature at Port Jervis, New York

$$\hat{\mu}(z) \approx 15.26 + 1.175 \cdot z$$

$$\log \hat{\sigma}(z) = 0.984 - 0.044 \cdot z$$

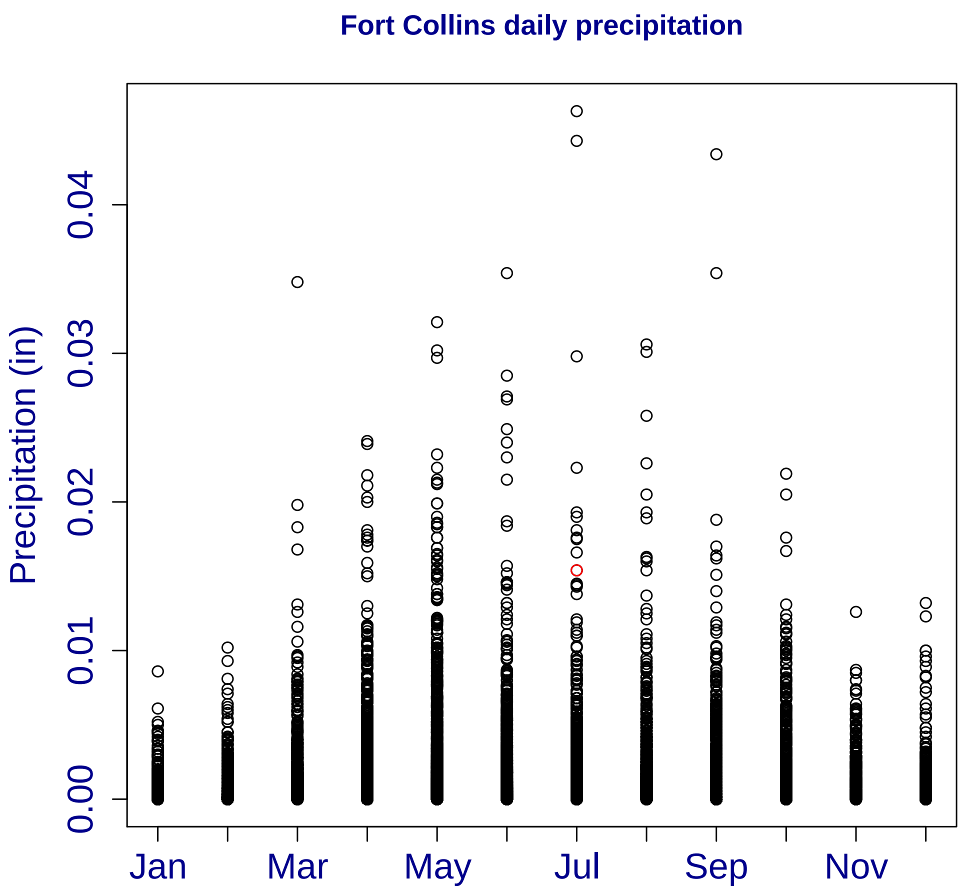
$$\xi(z) = -0.186$$

Likelihood ratio test for $\mu_1 = 0$ (p-value < 0.001)
Likelihood ratio test for $\sigma_1 = 0$ (p-value ≈ 0.635)

Non-Stationarity

Cyclic variation

Fort Collins, Colorado precipitation



Non-Stationarity

Cyclic variation

Fort Collins, Colorado precipitation Orthogonal approach. First fit annual cycle to Poisson rate parameter ($T = 365.25$):

$$\log \lambda(t) = \lambda_0 + \lambda_1 \sin\left(\frac{2\pi t}{T}\right) + \lambda_2 \cos\left(\frac{2\pi t}{T}\right)$$

Giving

$$\log \hat{\lambda}(t) \approx -3.72 + 0.22 \sin\left(\frac{2\pi t}{T}\right) - 0.85 \cos\left(\frac{2\pi t}{T}\right)$$

Likelihood ratio test for $\lambda_1 = \lambda_2 = 0$ (p-value ≈ 0).

Non-Stationarity

Cyclic variation

Fort Collins, Colorado precipitation Orthogonal approach. Next fit GPD with annual cycle in scale parameter.

$$\log \sigma^*(t) = \sigma_0^* + \sigma_1^* \sin\left(\frac{2\pi t}{T}\right) + \sigma_2^* \cos\left(\frac{2\pi t}{T}\right)$$

Giving

$$\log \hat{\sigma}^*(t) \approx -1.24 + 0.09 \sin\left(\frac{2\pi t}{T}\right) - 0.30 \cos\left(\frac{2\pi t}{T}\right)$$

Likelihood ratio test for $\sigma_1^* = \sigma *_2 = 0$ (p-value $< 10^{-5}$)

Non-Stationarity

Cyclic variation

Fort Collins, Colorado precipitation

Annual cycle in location and scale parameters of the GEV re-parameterization approach point process model with $t = 1, 2, \ldots,$ and $T = 365.25$.

$$\mu(t) = \mu_0 + \mu_1 \sin\left(\frac{2\pi t}{T}\right) + \mu_2 \cos\left(\frac{2\pi t}{T}\right)$$

$$\log \sigma(t) = \sigma_0 + \sigma_1 \sin\left(\frac{2\pi t}{T}\right) + \sigma_2 \cos\left(\frac{2\pi t}{T}\right)$$

$$\xi(t) = \xi$$

Non-Stationarity

Cyclic variation

Fort Collins, Colorado precipitation

$$\hat{\mu}(t) \approx 1.281 - 0.085 \sin\left(\frac{2\pi t}{T}\right) - 0.806 \cos\left(\frac{2\pi t}{T}\right)$$

$$\log \hat{\sigma}(t) \approx -0.847 - 0.123 \sin\left(\frac{2\pi t}{T}\right) - 0.602 \cos\left(\frac{2\pi t}{T}\right)$$

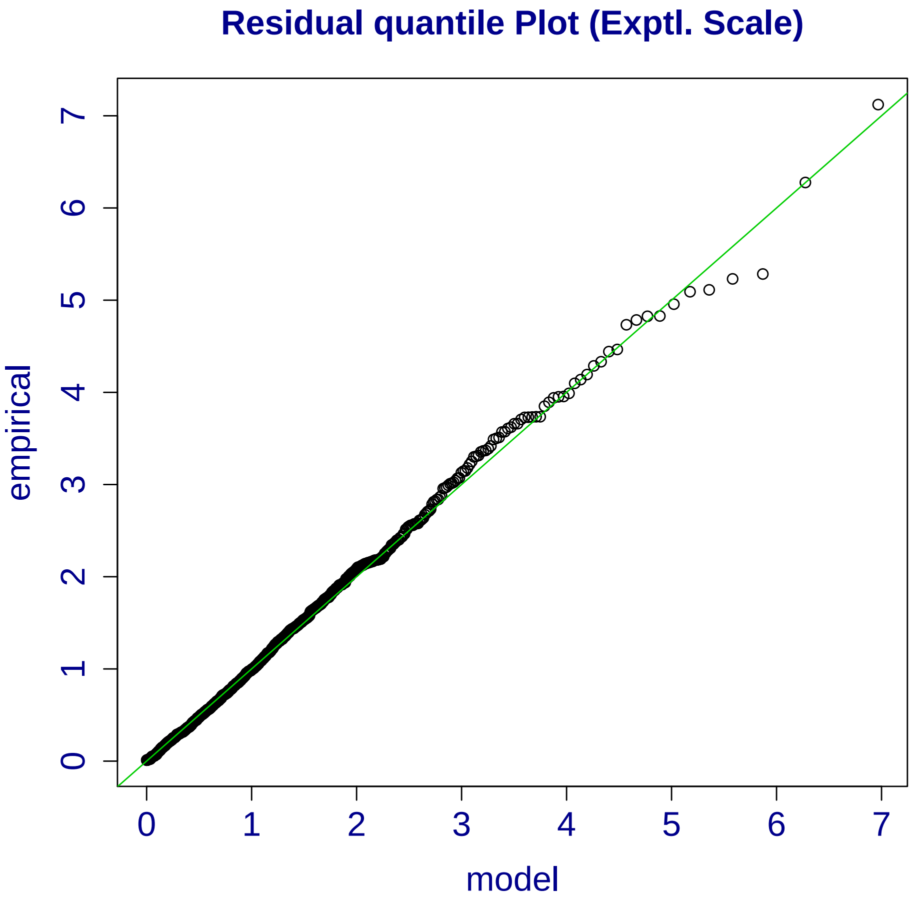
$$\hat{\xi} \approx 0.182$$

Likelihood ratio test for $\mu_1 = \mu_2 = 0$ (p-value ≈ 0).
Likelihood ratio test for $\sigma_1 = \sigma_2 = 0$ (p-value ≈ 0).

Non-Stationarity

Cyclic variation

Fort Collins, Colorado precipitation



Risk Communication (Under Non-Stationarity)

Return period/level does not make sense anymore because of changing distribution (e.g., with time). Often, one uses an "effective" return period/level instead. That is, compute several return levels for varying probabilities over time. Can also determine a single return period/level assuming temporal independence.

$$1 - \frac{1}{m} = \Pr\left\{\max(X_1, \ldots, X_n) \leq z_m\right\} \approx \prod_{i=1}^{n} p_i,$$

where

$$p_i = \begin{cases} 1 - \frac{1}{n} y_i^{-1/\xi_i} & , \text{ for } y_i > 0, \\ 1 & , \text{ otherwise} \end{cases}$$

where $y_i = 1 + \frac{\xi_i}{\sigma_i}(z_m - \mu_i)$, and (μ_i, σ_i, ξ_i) are the parametrs of the point process model for observation i. Can be easily solved for z_m (using numerical methods). Difficulty is in calculating the uncertainty (See Coles, 2001, chapter 7).

Heat Waves/Hot Spells

Long stretches of high (but not necessarily extreme) temperatures without relief can have devastating impacts.

- EVA may not be needed here.
- Point process approach may be useful.

Short stretches of high temperatures accompanied with an extremely hot day can also have devastating impacts.

- EVA may be useful here (particularly point process approach).
- Need more information than just a block extreme or threshold excess.

References of papers using EVA to analyze weather spells can be found at Rick Katz' *Extremes* web page:
http://www.isse.ucar.edu/extremevalues/biblio.html#spells

Heat Waves/Hot Spells

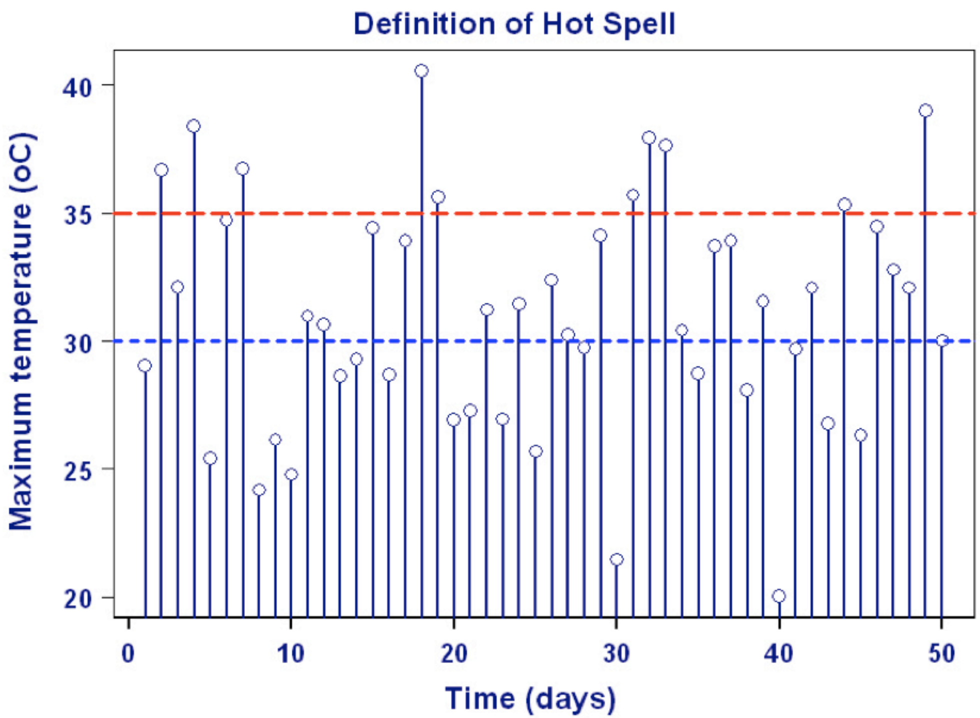


Image from: Katz, R.W., E.M. Furrer, and M.D. Walter, 2009: Statistical modeling of hot spells and heat waves. *International Conference on Extreme Value Analysis*, Fort Collins, CO.

The R programming language

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.r-project.org

Vance A, 2009. Data analysts captivated by R's power. *New York Times*, 6 January 2009. Available at: http://www.nytimes.com/2009/01/07/technology/business-computing/07program.html?_r=2

R preliminaries

Assuming R is installed on your computer...

In linux, unix, and Mac (terminal/xterm) the directory in which R is opened is (by default) the current working directory. In Windows (Mac GUI?), the working directory is usually in one spot, but can be changed (tricky).

Open an R workspace:
Type `R` at the command prompt (linux/unix, Mac terminal/xterm) or double click on R's icon (Windows, Mac GUI).

`getwd()` # Find out which directory is the current working directory.

R preliminaries

Assigning vectors and matrices to objects:

```
# Assign a vector containing the numbers -1, 4
# and 0 to an object called 'x'
x <- c( -1, 4, 0)

# Assign a 3 × 2 matrix with column vectors:  2, 1, 5 and
# 3, 7, 9 to an object called 'y'.
y <- cbind( c( 2, 1, 5), c(3, 7, 9))

# Write 'x' and 'y' out to the screen.
x
y
```

R preliminaries

Saving a workspace and exiting

```
# To save a workspace without exiting R.
save.image()

# To exit R while also saving the workspace.
q("yes")

# Exit R without saving the workspace.
q("no")

# Or, interactively...
q()
```

R preliminaries

Subsetting vectors:

```
# Look at only the 3-rd element of 'x'.
x[3]

# Look at the first two elements of 'x'.
x[1:2]

# The first and third.
x[c(1,3)]

# Everything but the second element.
x[-2]
```

R preliminaries

Subsetting matrices:

```
# Look at the first row of 'y'.
y[1,]

# Assign the first column of 'y' to a vector called 'y1'.
# Similarly for the 2nd column.
y1 <- y[,1]
y2 <- y[,2]

# Assign a "missing value" to the second row, first column
# element of 'y'.
y[2,1] <- NA
```

R preliminaries

Logicals and Missing Values:

```
# Do 'x' and/or 'y' have any missing values?
any( is.na( x))
any( is.na( y))

# Replace any missing values in 'y' with -999.0.
y[ is.na( y)] <- -999.0

# Which elements of 'x' are equal to 0?
x == 0
```

R preliminaries

Contributed packages

```
# Install some useful packages.  Need only do once.
install.packages( c("fields",       # A spatial stats package.
                    "evd",          # An EVA package.
                    "evdbayes",     # Bayesian EVA package.
                    "ismev",        # Another EVA package.
                    "maps",         # For adding maps to plots.
                    "SpatialExtremes"))

# Now load them into R. Must do for each new session.
library( fields)
library( evd)
library( evdbayes)
library( ismev)
library( SpatialExtremes)
```

R preliminaries

See hierarchy of loaded packages:

```
search()
```

```
# Detach the 'SpatialExtremes' package.
detach(pos=2)
```

See how to reference a contributed package:

```
citation("fields")
```

R preliminaries

Help files

Getting help from a package or a function

```
help( ismev)
```

Alternatively, can use `?`. For example,

```
?extRemes
```

For functions,

```
?gev.fit
```

Example data sets:

```
?HEAT
```

R preliminaries

Basics of plotting in R:

- First must open a device on which to plot.
  - Most plotting commands (e.g., `plot`) open a device (that you can see) if one is not already open. If a device is open, it will write over the current plot.
  - `X11()` will also open a device that you can see.
  - To create a file with the plot(s), use `postscript`, `jpeg`, `png`, or `pdf` (before calling the plotting routines. Use `dev.off()` to close the device and create the file.
- `plot` and many other plotting functions use the `par` values to define various characteristics (e.g., margins, plotting symbols, character sizes, etc.). Type `help( plot)` and `help( par)` for more information.

R preliminaries

Simple plot example.

```
plot( 1:10, z <- rnorm(10), type="l", xlab="", ylab="z",
           main="Std Normal Random Sample")
points( 1:10, z, col="red", pch="s", cex=2)
lines( 1:10, rnorm(10), col="blue", lwd=2, lty=2)

# Make a standard normal qq-plot of 'z'.
qqnorm( z)

# Shut off the device.
dev.off()
```

References

Coles S, 2001. An introduction to statistical modeling of extreme values. Springer, London. 208 pp.

Katz RW, MB Parlange, and P Naveau, 2002. Statistics of extremes in hydrology. *Adv. Water Resources*, **25**:1287–1304.

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