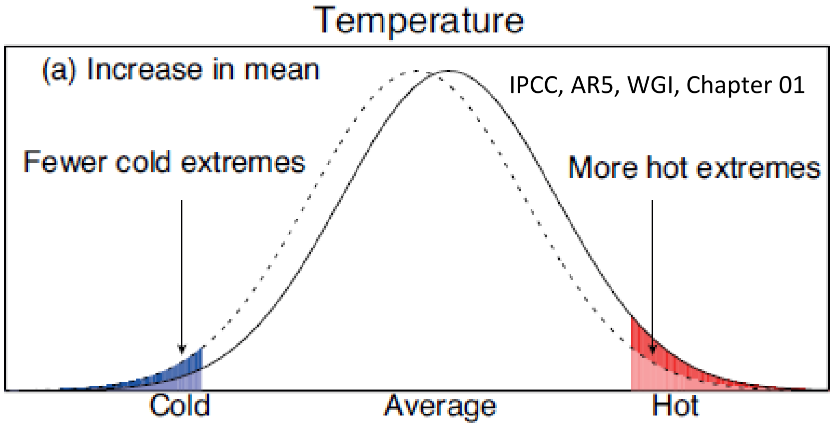
Extreme Value Theory: A primer

Harald E. Rieder
Lamont-Doherty Earth Observatory
9/8/2014

Based on :
Coles (2001) An Introduction to Statistical Modelling of Extreme Values, Springer
Davison (2005): Extreme Values, Encyclopedia of Biostatistics, Wiley.
Katz (2013): Chapter 2 - Statistical Methods for Nonstationary Extremes, in Extremes in a Changing Climate, Detection, Analysis and Uncertainty, Springer
Cooley (2013): Chapter 4 - Return Periods and Return Levels under Climate Change, in Extremes in a Changing Climate, Detection, Analysis and Uncertainty, Springer

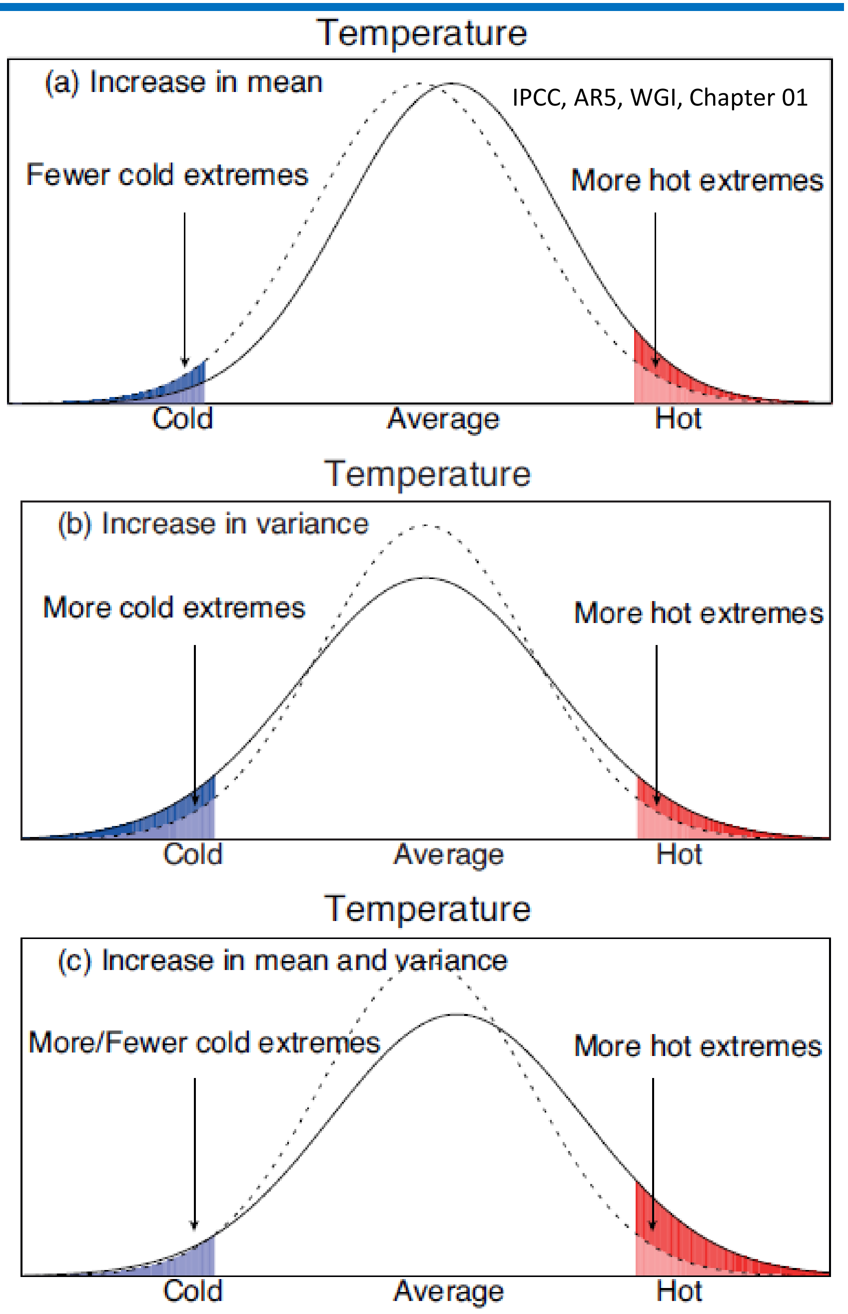
Introduction

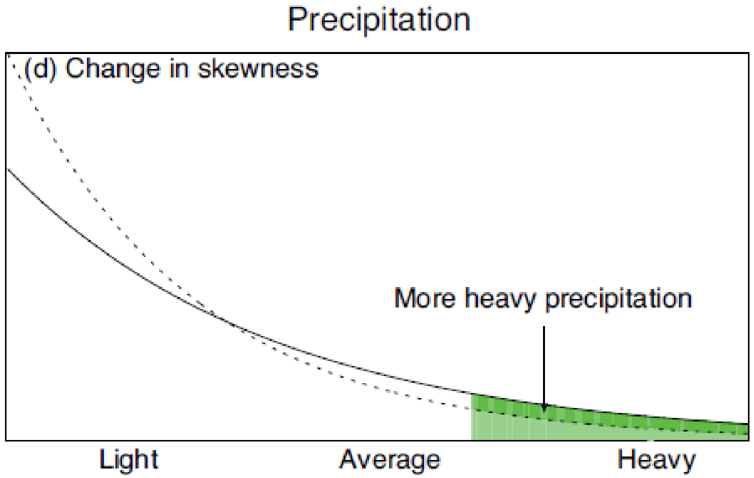
The probability of occurrence of values of a climate or weather variable can be described by a probability density function (PDF) that for some variables (e.g., temperature) is shaped similar to a Gaussian curve. A PDF is a function that indicates the relative chances of occurrence of different outcomes of a variable. Simple statistical reasoning indicates that substantial changes in the frequency of extreme events (e.g., the maximum possible 24-hour rainfall at a specific location) can result from a relatively small shift in the distribution of a weather or climate variable. Figure 1.8a shows a schematic of such a PDF and illustrates the effect of a small shift in the mean of a variable on the frequency of extremes at either end of the distribution. An increase in the frequency of one extreme (e.g., the number of hot days) can be accompanied by



Introduction

The probability of occurrence of values of a climate or weather variable can be described by a probability density function (PDF) that for some variables (e.g., temperature) is shaped similar to a Gaussian curve. A PDF is a function that indicates the relative chances of occurrence of different outcomes of a variable. Simple statistical reasoning indicates that substantial changes in the frequency of extreme events (e.g., the maximum possible 24-hour rainfall at a specific location) can result from a relatively small shift in the distribution of a weather or climate variable. Figure 1.8a shows a schematic of such a PDF and illustrates the effect of a small shift in the mean of a variable on the frequency of extremes at either end of the distribution. An increase in the frequency of one extreme (e.g., the number of hot days) can be accompanied by a decline in the opposite extreme (in this case the number of cold days such as frost days). Changes in the variability, skewness or the shape of the distribution can complicate this simple picture (Figure 1.8b, c and d).

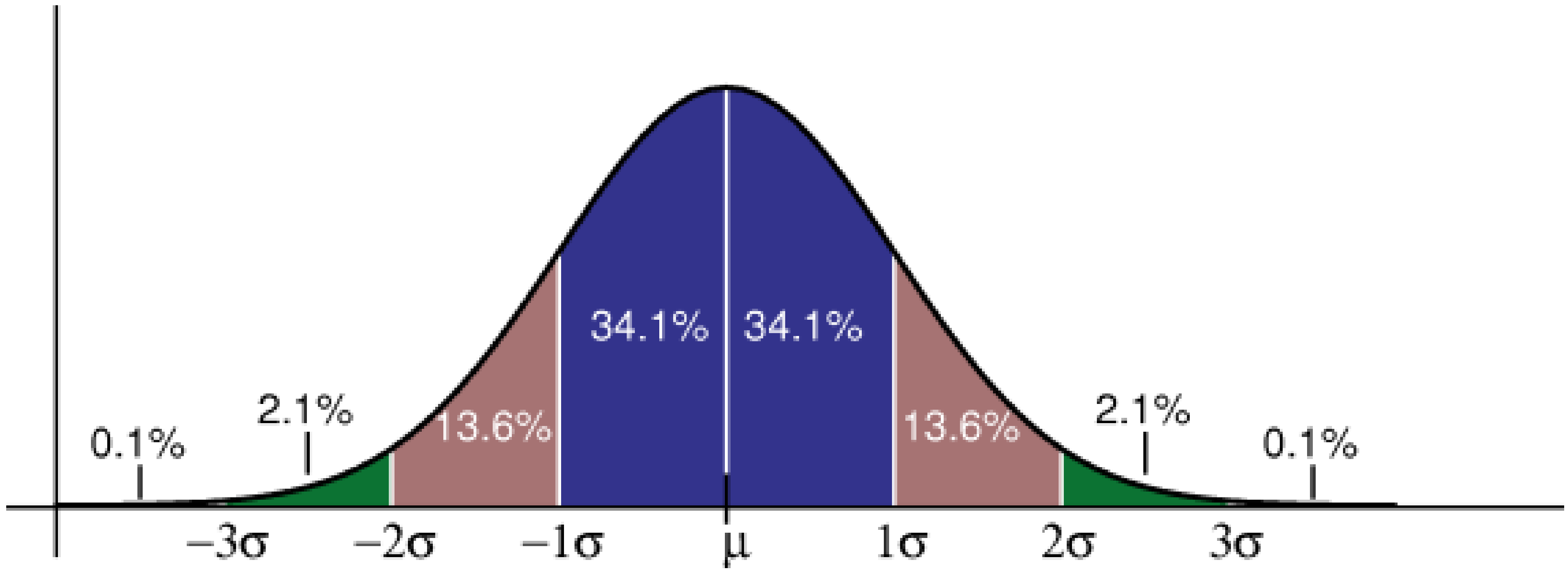




Normal Distribution (or Gaussian, or 'bell curve')
is a continuous probability distribution given by

$$F(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where the parameter μ is the mean of the distribution (and also its median and mode) and the parameter σ is the standard deviation.



Source: introcs.cs.princeton.edu

Statistical extreme value theory is a field of statistics dealing with extreme values, i.e., large deviations from the median of probability distributions. The theory assesses the type of probability distribution generated by processes.

Extreme value distributions are the limiting distributions for the minimum or the maximum of large collections of independent random variables from the same arbitrary distribution. By definition extreme value theory focuses on limiting distributions (which are distinct from the normal distribution).

Two approaches exist for practical extreme value applications. The first method relies on deriving **block maxima** (minima) series, the second method relies on extracting **peak values above (below) a certain threshold** from a continuous record.

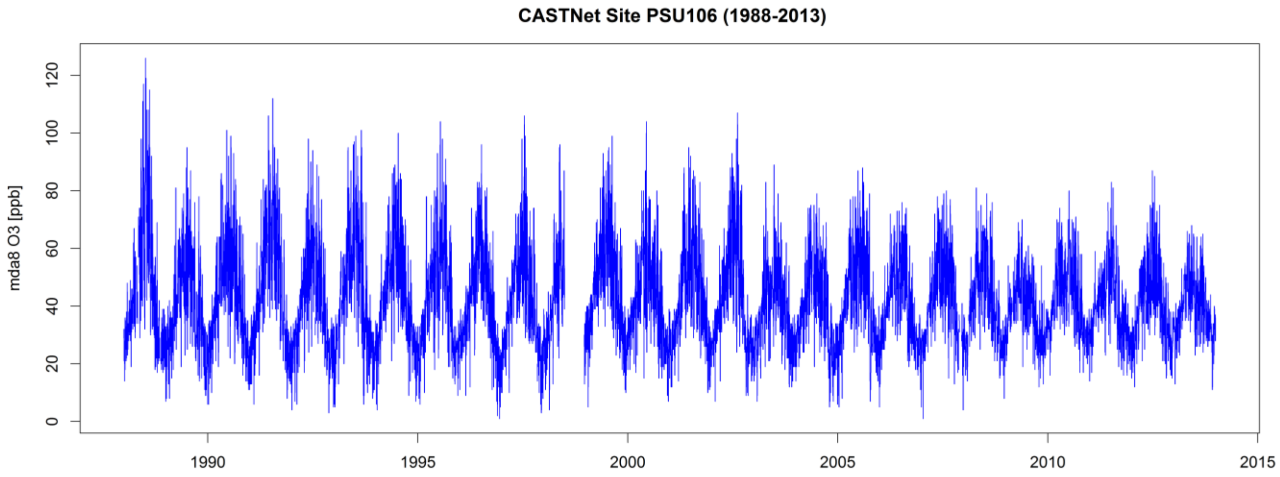
A third approach the so-called **r-largest order statistics** represents a compromise between the block maxima and peak over threshold approach.

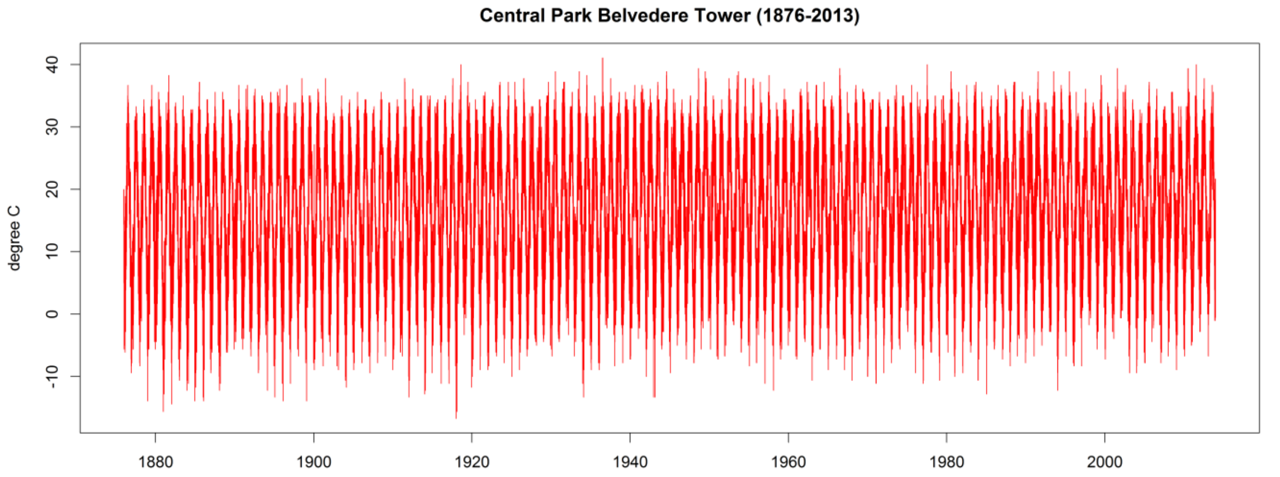
Data Sets

Let's look on some examples with real world data:

(1) maximum daily 8-hour surface ozone from CastNet Site PSU106

(2) daily maximum temperature from NYC Central Park Belvedere Tower





Data Sets

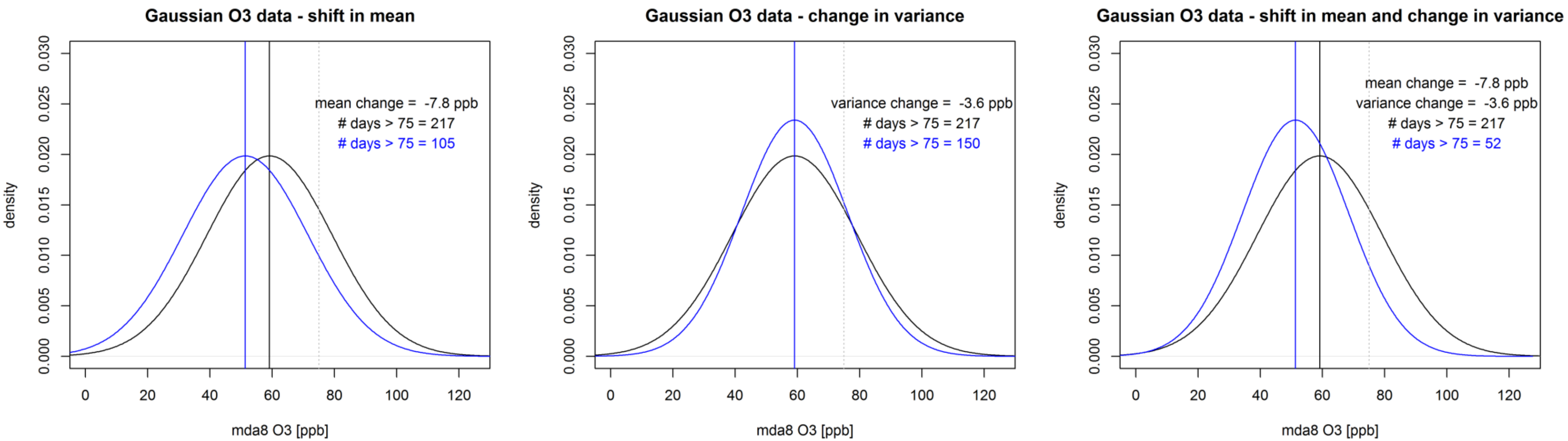
For simplicity we focus on summer time (JJA) data only
and we consider extreme values as:

(1) mda8 O3 > 75 ppb (NAAQS)
1988-2000 vs 2001-2013: shift in mean -7.8 ppb; change in variance -3.6 ppb

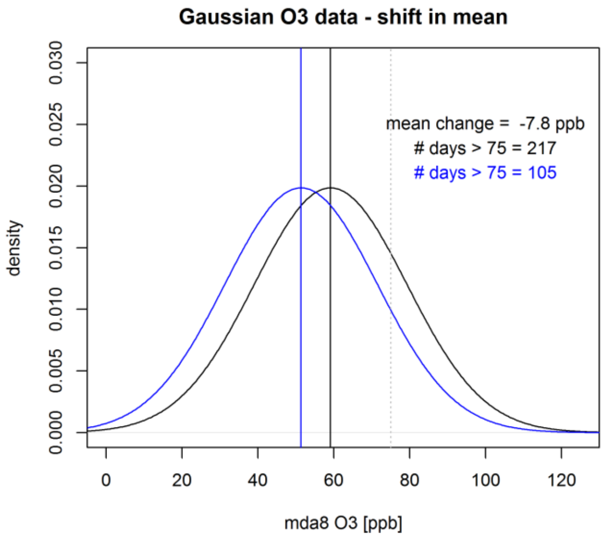
(2) Tmax ≥ 25 degree C (summer day)
1876-1944 vs 1945-2013: shift in mean +0.86 C; change in variance -0.025 C

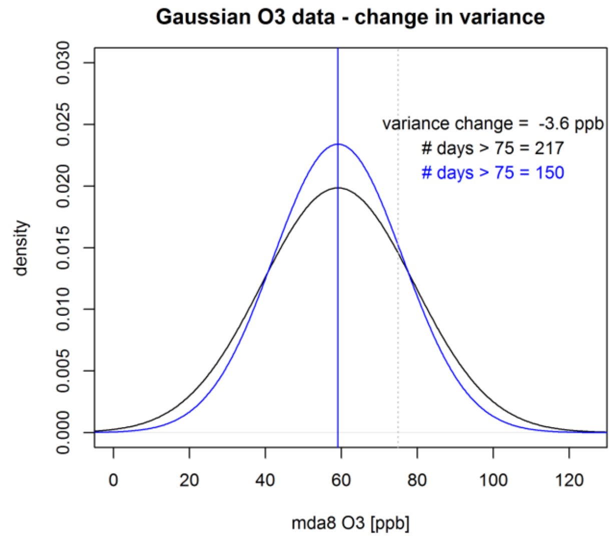
We want to visualize how these changes in mean or variance or both affect the distributions and in particular the probabilistic frequency of extremes.

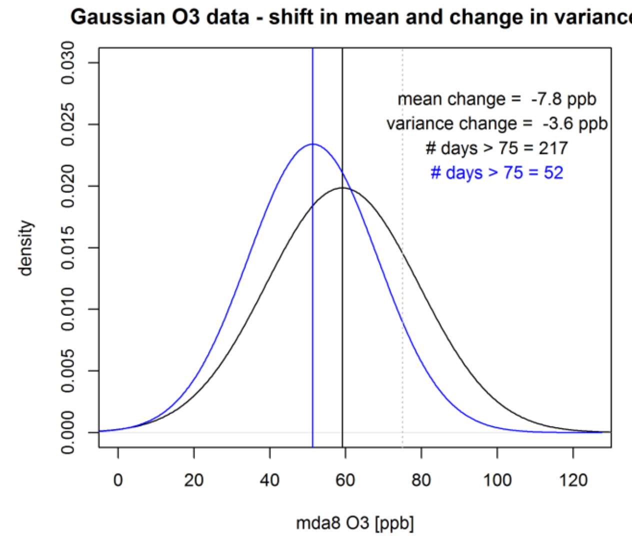
Influence of shift in mean and/or change in variance

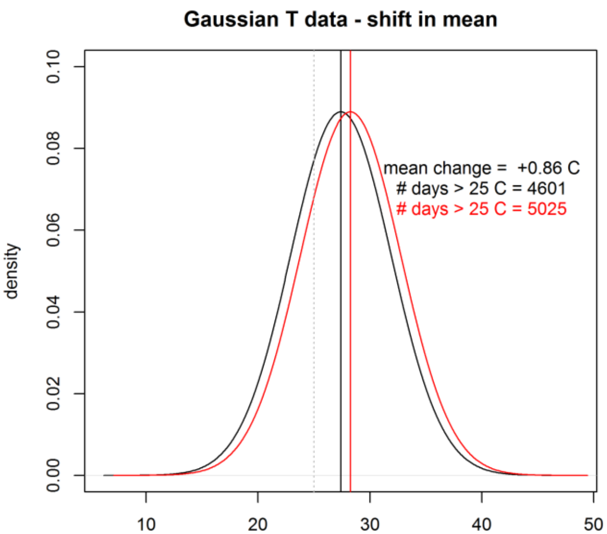


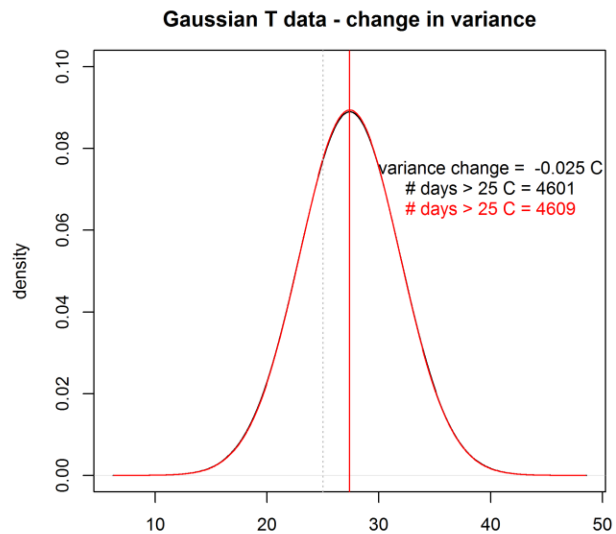
Influence of shift in mean and/or change in variance

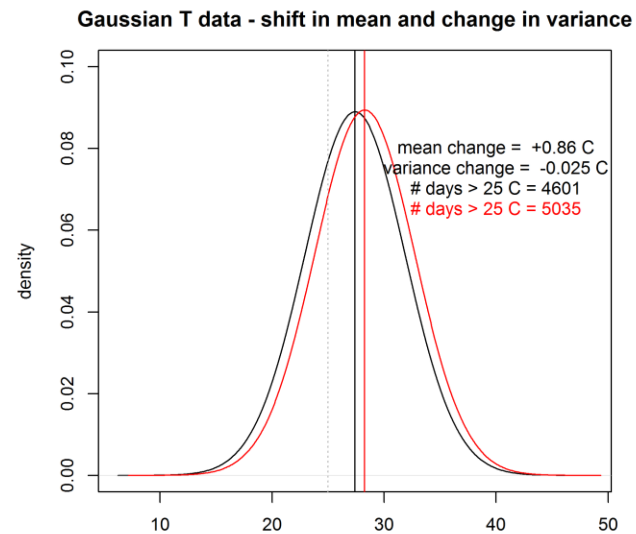






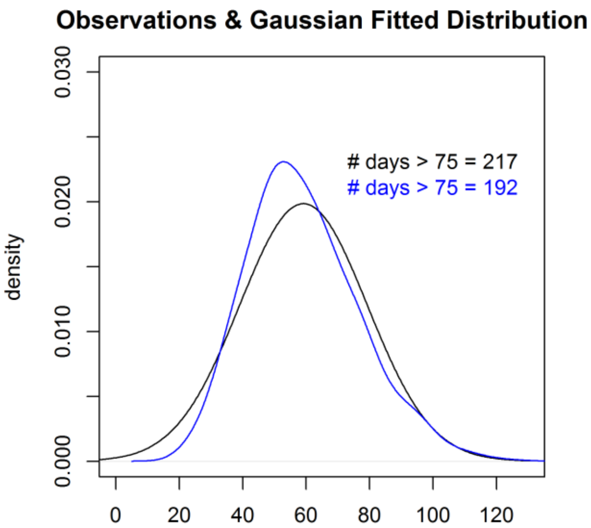


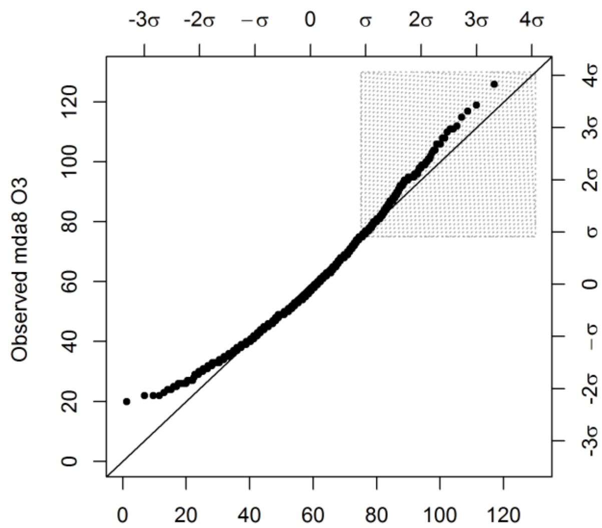




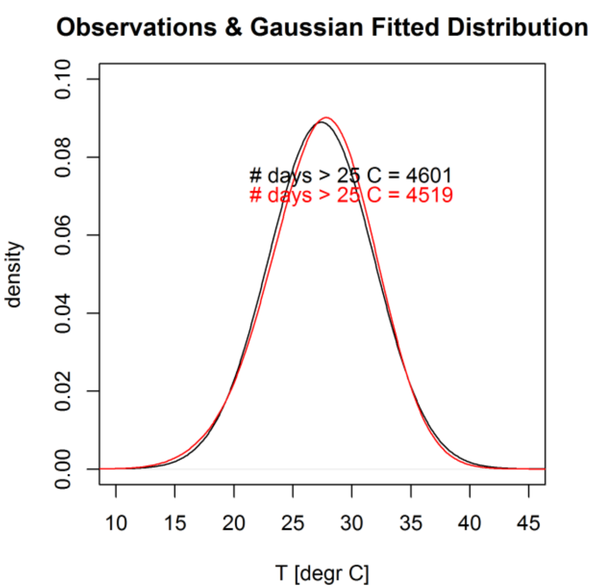
Comparison of observed distributions with least-square fitted normal distributions

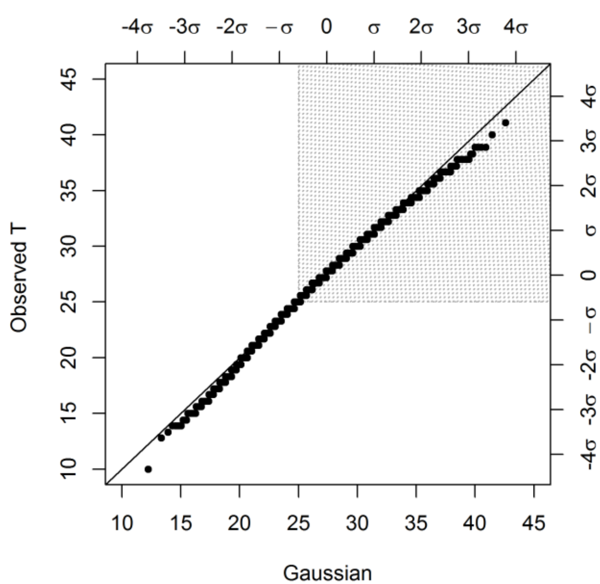
Compare observed distributions with Normal distributions





Quantile	Observed	Gaussian
0.10	38 ppb	36 ppb
0.75	70 ppb	71 ppb
0.9	82 ppb	81 ppb
0.95	92 ppb	88 ppb
0.99	106 ppb	100 ppb





Quantile	Observed	Gaussian
0.10	22.2 C	22.3 C
0.75	30.0 C	30.1 C
0.9	32.2 C	32.5 C
0.95	33.9 C	34.0 C
0.99	36.1 C	36.7 C

Testing for normality

The Shapiro-Wilk test is a standard test to check whether a sample $X_1,\ldots,X_k$ stems from a normally distributed population.

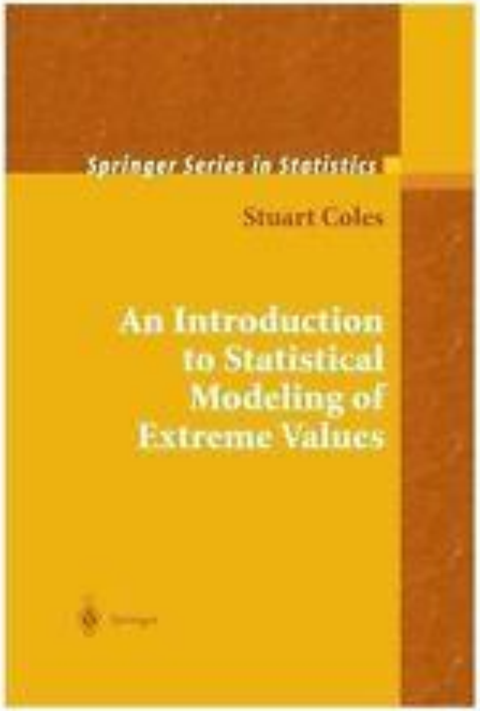
The test statistic (W) is $W = \frac{(\sum_{i=1}^{k} a_i x_{(i)})^2}{\sum_{i=1}^{k}(x_i - \bar{x})^2}$

where x_i is the *i*th-smallest number in the sample, $\bar{x} = (x_i + \cdots + x_k)/k$ is the sample mean, and a_i are constants $(a_1, \ldots, a_k) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}}$, where $m = (m_1, \ldots, m_k)^T$ and $m_1, \ldots, m_k$ are the iid expected values of the order statistics from the normal distribution and V represents the covariance matrix of these order statistics.

The **null hypothesis, the population is normally distributed**, is rejected when the p-value of the test is below a defined significance value (e.g., 0.05)

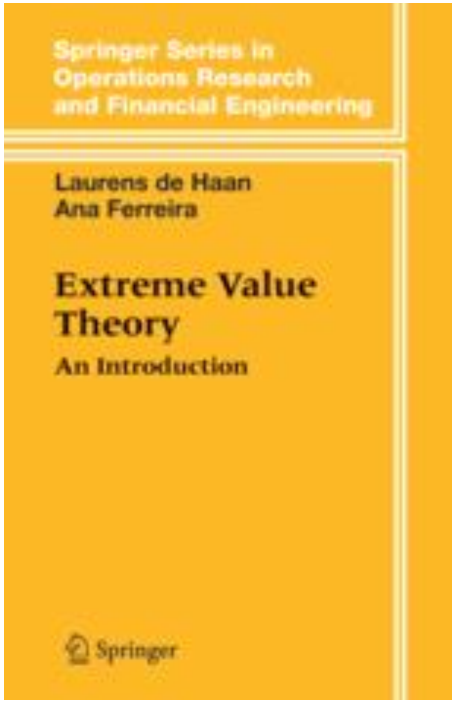
Extreme value theory (EVT) is concerned with the occurrence and sizes of rare events, be they larger or smaller than usual.

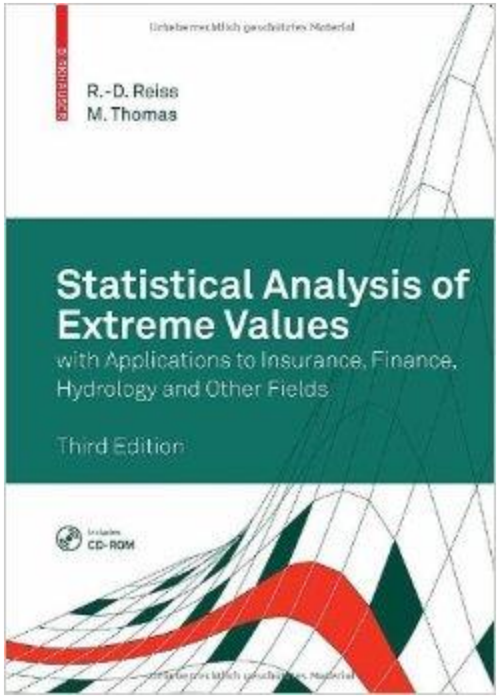
Here we want to review briefly the most common EVT approaches and models and look into some applications.

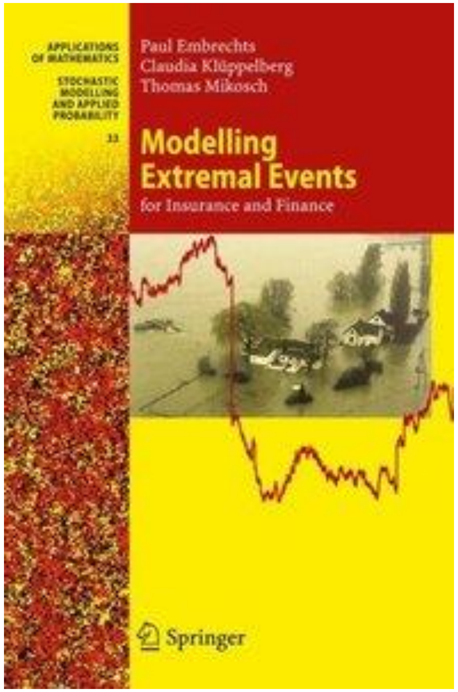


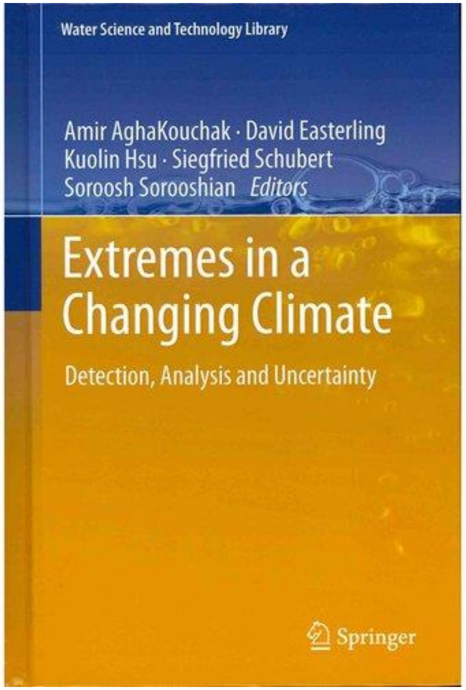
There has been rapid development over the last decades in both theory and applications. A comprehensive introduction to statistics of extremes is provided by Coles (2001).

Further Reading









Frequently discussion of extremes concerns high extremes, maxima. Also we will focus our initial discussion on maxima. Though it shall be noted that dealing with minima follows the same approaches and in applications all needed to be done is reverse the sings of the observations and apply procedures for maxima as

$$\min(x_i) = -\max(-x_i)$$

Block Maxima

The **Extremal Types Theorem (ETT)** (e.g. Leadbetter et al., 1983) addresses the following question: Given a set of independent identically distributed random variables $X_1, \ldots, X_k$, what are the possible limiting distributions of

$$M_k = a_k[\max(X_1, \ldots, X_k) - b_k] \; as \; k \to \infty \, ?$$

$$F^k\left(\frac{x - b_k}{a_k}\right) \xrightarrow[k\to\infty]{} G(x)$$

The answer is that if a nondegenerate limiting cumulative distribution (cdf) exists for some sequences of constants a_k and b_k, it must fall into one of the three classes

$$\text{I: } F(x) = \exp[-e^{-x}], \quad -\infty < x < \infty,$$

$$\text{II: } F(x) = \begin{cases} 0, & x \leq 0, \\ \exp(-x^{-\alpha}), & x > 0, \alpha > 0, \end{cases}$$

$$\text{III: } F(x) = \begin{cases} \exp[-(-x)^{\alpha}], & x < 0, \alpha > 0, \\ 1, & x \geq 0. \end{cases}$$

$$\text{I: } F(x) = \exp[-e^{-x}], \quad -\infty < x < \infty,$$

$$\text{II: } F(x) = \begin{cases} 0, & x \leq 0, \\ \exp(-x^{-\alpha}), & x > 0, \alpha > 0, \end{cases}$$

$$\text{III: } F(x) = \begin{cases} \exp[-(-x)^{\alpha}], & x < 0, \alpha > 0, \\ 1, & x \geq 0. \end{cases}$$

The three types of distributions represent the Gumbel, Frechet and Weibull distributions. **The ETT guarantees that if a limit exists for maxima, it must have one of these specified forms**.

Generalized Extreme Value Distribution

In a more modern approach these distributions are combined into the **generalized extreme value distribution (GEV)** with cdf

$$H(y) = \exp\left\{-\left[1+\xi\left(\frac{y-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, -\infty < \mu, \xi < \infty, \sigma > 0,$$

define for values of y for which $1+ \xi(y - \mu)/\sigma > 0$.

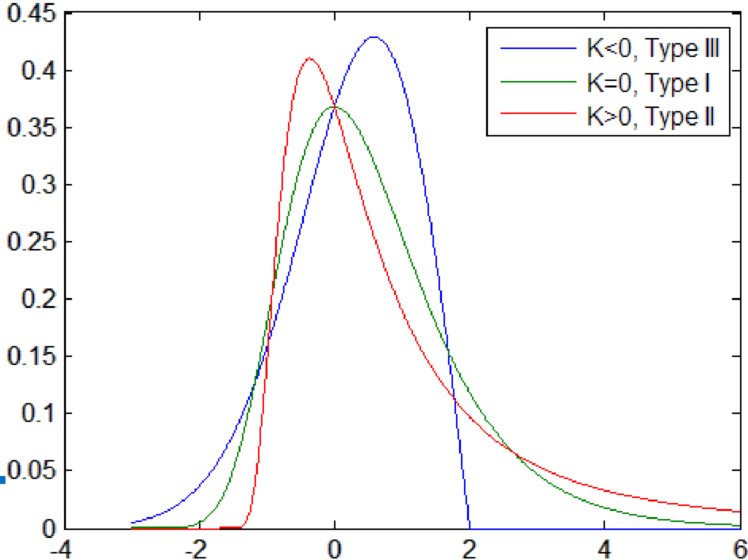
where μ is the location parameter, ξ is the shape parameter, and $\sigma > 0$ is the scale parameter.

The shape parameter ξ governs the distribution type:

type I with $\xi = 0$ (Gumbel,light tailed)

type II with $\xi > 0$ (Frechet, heavy tailed)

type III with $\xi < 0$ (Weibull, bounded)



GEV type I with $\xi = 0$ (Gumbel, light tailed)

Domain of attraction for many common distributions (e.g., normal, exponential, gamma), not frequently found to fit 'real world data'

GEV type II with $\xi > 0$ (Frechet, heavy tailed)

Fits found for precipitation, stream flow, economic damage

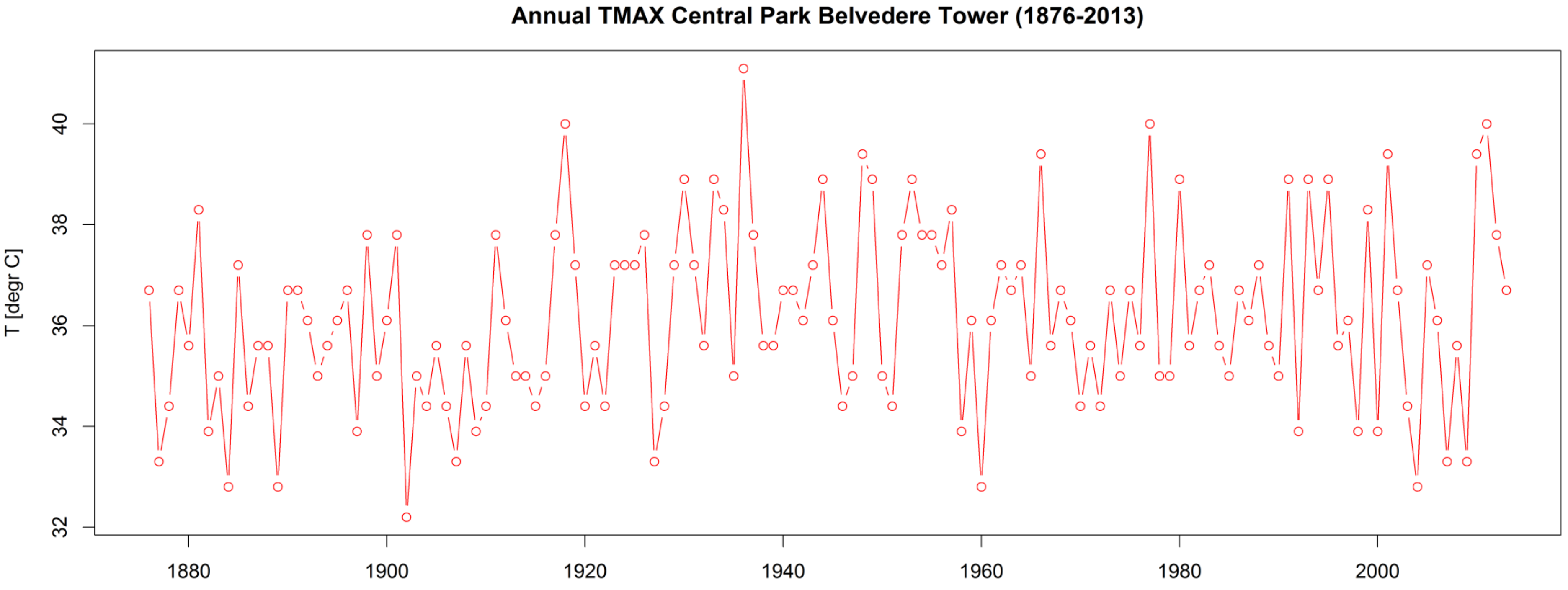
GEV type III with $\xi < 0$ (Weibull, bounded)

Fits found for temperature, wind speed, pollutants, sea level

Block Maxima - Application

It is important to note that the **location parameter μ is not the mean** but does represent the 'center' of the distribution, and the **scale parameter σ is not the standard deviation** but does govern the size of the deviations about μ.

A typical application would be to fit a GEV to the annual maximum of a variable. Note that the block size for maxima is freely variable though applications must be consistent with the maxima of a given block.



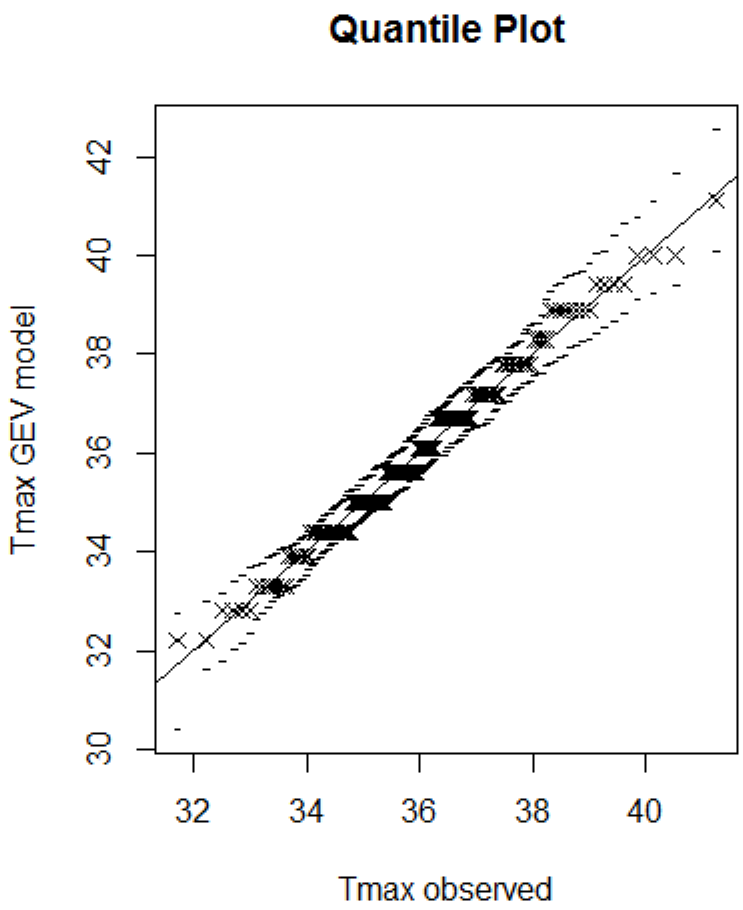
Fit GEV distribution to annual MAX of summertime (JJA) temperature

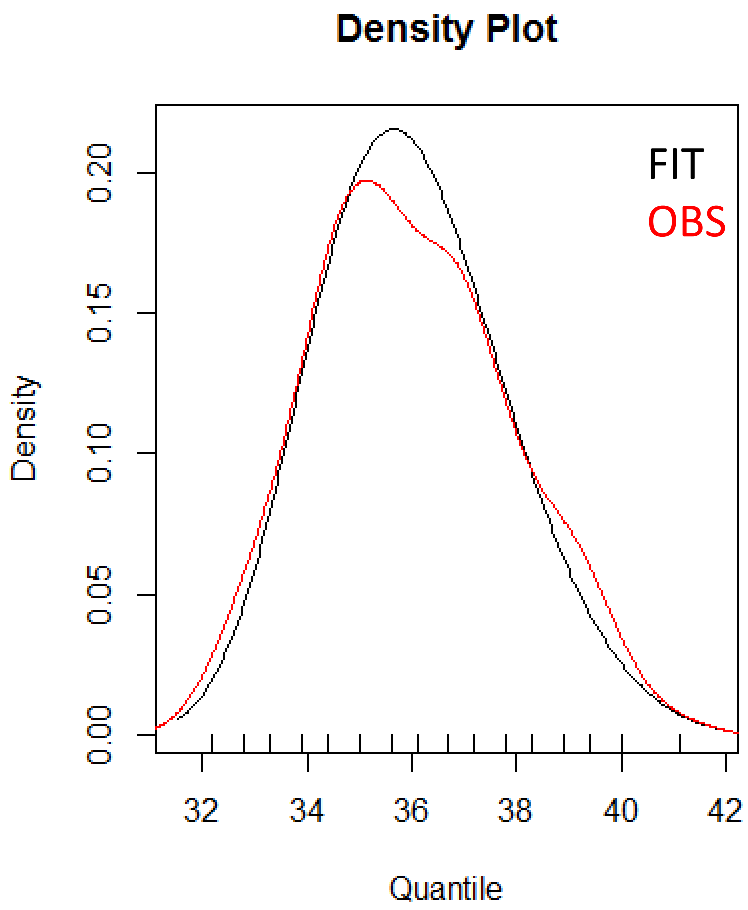
$$H(y) = \exp\left\{-\left[1+\xi\left(\frac{y-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, -\infty < \mu, \xi < \infty, \sigma > 0,$$

μ = 35.28 ($\pm$ 0.16)
σ = 1.74 ($\pm$ 0.12)
ξ = -0.19 ($\pm$ 0.06)

GEV distribution type III with $\xi < 0$ (Weibull, bounded)

Block Maxima - Application

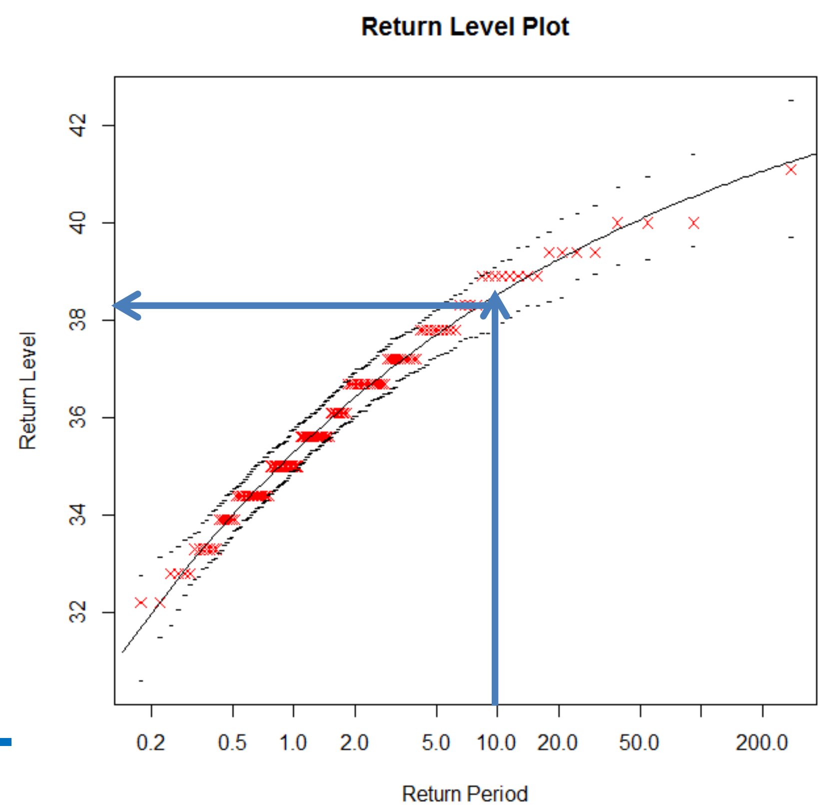




Return Levels

The fitted distribution than can be used to estimate the **m-year return level**, which represents the high quantile for which the probability that the annual maximum exceeds this quantile is $1/m$.

Under the assumption of stationarity the return level is the same for all years, giving rise to the notion of the return period. The **return period of a particular event is the inverse of the probability that the event will be exceeded in any given year**, i.e. the m-year return level is associated with a return period of m years.



Return Levels

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

Computing the return level r_m such that

$$\text{GEV}(r_m) = 1 - m$$

$$r_m = \text{GEV}^{-1}(1 - m)$$

Hence, $r_m = \mu + \frac{\sigma}{\xi}\left([-\ln(1-m)]^{-\xi} - 1\right)$

$$r_m = \mu + \frac{\sigma}{\xi}\left([-\ln(1-m)]^{-\xi} - 1\right)$$

Estimating the return level r_m

$$r_m = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}\left([-\ln(1-m)]^{-\hat{\xi}} - 1\right)$$

Where GEV parameter estimates $\hat{\mu}, \hat{\sigma}, \hat{\xi}$, are derived via

-Maximum likelihood estimation (MLE) (inbuilt in gev packages in R, MatLab)
-Methods of moments (PWM, GPWM)
-Exhaustive tail-index approaches

In the stationary case there is a one to one relationship between the m-year return level and the m-year return period (reciprocal of exceedance in any given year).

A 10-year event might be more interpretable by lay audiences as a 0.1 probability of occurrence in any given year. Though this leads to two possible interpretations of the 'm-year event'.

(1) The expected number of events in m years is 1.
(2) The expected waiting time till the next exceedance is m years.

Under the assumption of stationarity both of this interpretations are correct.

We will discuss return level interpretations under non-stationarity a bit later in this class.

Extending beyond block maxima

One argument against the application of a block-maximum approach is that use of maxima alone is wasteful of data: most of the information in the sample is ignored.

To overcome this limitation several other approaches have been developed based on the **point process characterization**.

Here we want to review 2 common applications:

(1) the r-largest order model for extremes

(2) peak over threshold modeling

r-largest order extremes

The r-largest observations among $Y_1,\ldots,Y_k$, will contain more information about the extremes than the maximum alone.

Let $y_t^{(1)} \geq \ \ldots \ y_t^{(r)}$ be the r-largest observation in a block, or equivalently time period, $t \in \{1,2,\ldots T\}$ and define the location μ, shape σ and scale ξ parameters as in the GEV then the likelihood L_t for block t based on the r-largest order statistics model is

$$L_t(\mu,\sigma,\xi) = \exp\left(-\left(1+\xi\frac{y_t^{(r)}-\mu}{\sigma}\right)_+^{-1/\xi}\right) \qquad (2)$$
$$\times \prod_{j=1}^{r}\frac{1}{\sigma}\left(1+\xi\frac{y_t^{(j)}-\mu}{\sigma}\right)_+^{-1/\xi-1}.$$

This expression is related to the GEV distribution whose probability density function is obtained on setting r =1.

r-largest order statistics application

Block maximum

μ = 35.28 ($\pm$ 0.16)
σ = 1.74 ($\pm$ 0.12)
ξ = -0.19 ($\pm$ 0.06)

10yr return level: 38.5
25yr return level: 39.4
50yr return level: 40.0
100yr return level: 40.6

3-largest observations per summer

μ = 34.69 ($\pm$ 0.11)
σ = 1.67 ($\pm$ 0.08)
ξ = -0.28 ($\pm$ 0.05)

10yr return level: 37.9
25yr return level: 38.8
50yr return level: 39.4
100yr return level: 39.8

The **peak over threshold approach** is based on the idea of modelling data over a high enough threshold.

The cdf of the amount by which an observation exceeds a high threshold u, given that it has done so, is

$$pr(X > u + y \mid X > u) = G(y) = 1 - \left(1 + \xi \frac{y - u}{\sigma}\right)^{-\frac{1}{\xi}},$$

$$\sigma > 0, \qquad y > u, \qquad 1 + \xi \frac{y - u}{\sigma} > 0;$$

which is called the **Generalized Pareto distribution**.

The shape parameter ξ has the same meaning as in the GEV type with

type I with $\xi = 0$ (light tailed, exponential type)

type II with $\xi > 0$ (heavy tailed, Pareto type)

type III with $\xi < 0$ (bounded, beta type)

It shall be noted that in the GPD setting the **scale parameter $\boldsymbol{\sigma}$ is dependent on the threshold.**

Meaning if the distribution of the excess Y_i has an exact GP distribution (rather than only approximate) then increasing the threshold from u to u^* would result in another GP distribution with the same shape parameter ξ, but an adjusted scale parameter given by

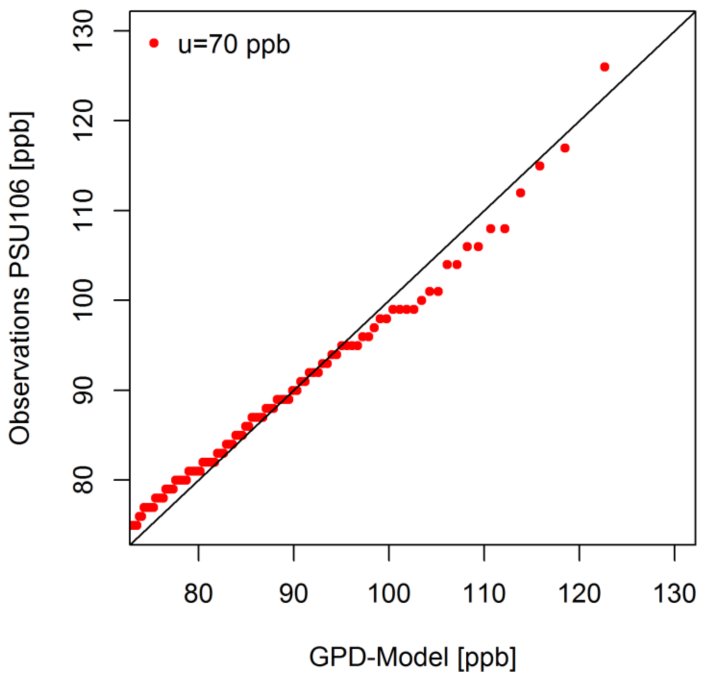
$$\sigma(u^*) = \sigma(u) + \xi(u^* - u), u^* > u$$

Note that the scale parameter would increase if $\xi > 0$ and decrease if $\xi < 0$. Consistent with the exponential distribution, there would be no change in the scale parameter if $\xi = 0$.

Selection of a threshold involves a delicate trade-off between bias and variance.

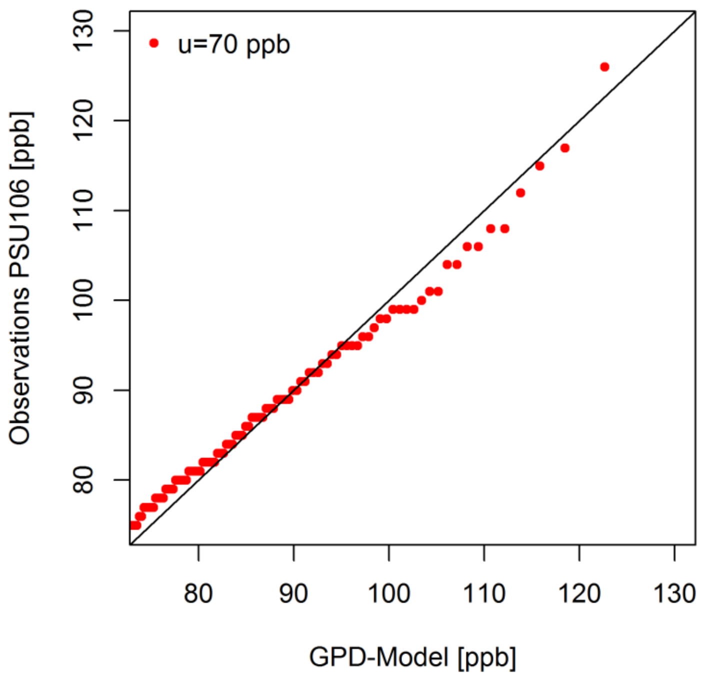
Too high a threshold will reduce the number of exceedances and thus increase the estimation variance and the reliability of the parameter estimates, whereas too low a threshold will induce a bias because the GPD will fit the exceedances poorly.

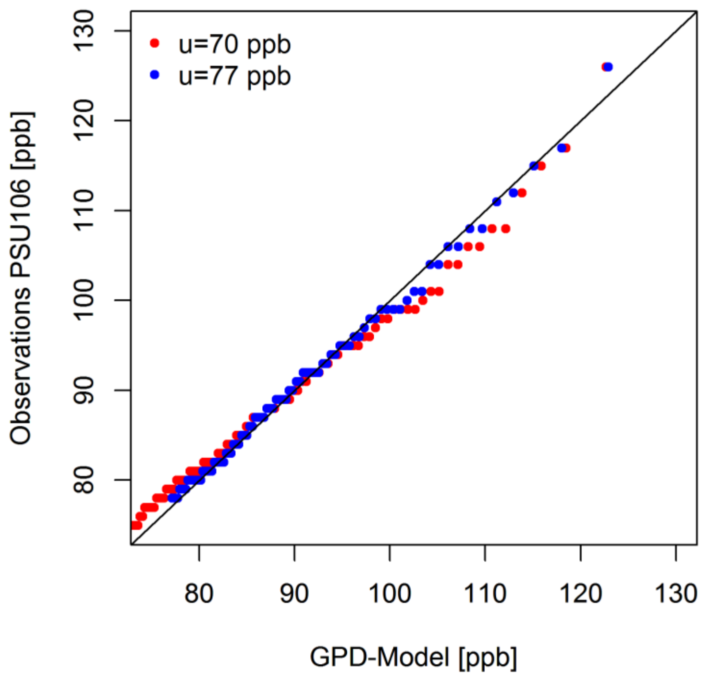
Peak over threshold - Application



Parameter	Model 1
μ	70
σ	22.77 (±2.48)
ξ	-0.37 (± 0.06)
AIC	**791.871**

Peak over threshold - Application





Parameter	Model 1	Model 2
μ	70	77
σ	22.77 (±2.48)	17.36 (±2.26)
ξ	-0.37 (± 0.06)	-0.29 (± 0.08)
AIC	**791.871**	**616.086**

$$\text{AIC} = 2k - 2\ln(L)$$

k … number of parameters

L … maximized value of likelihood function

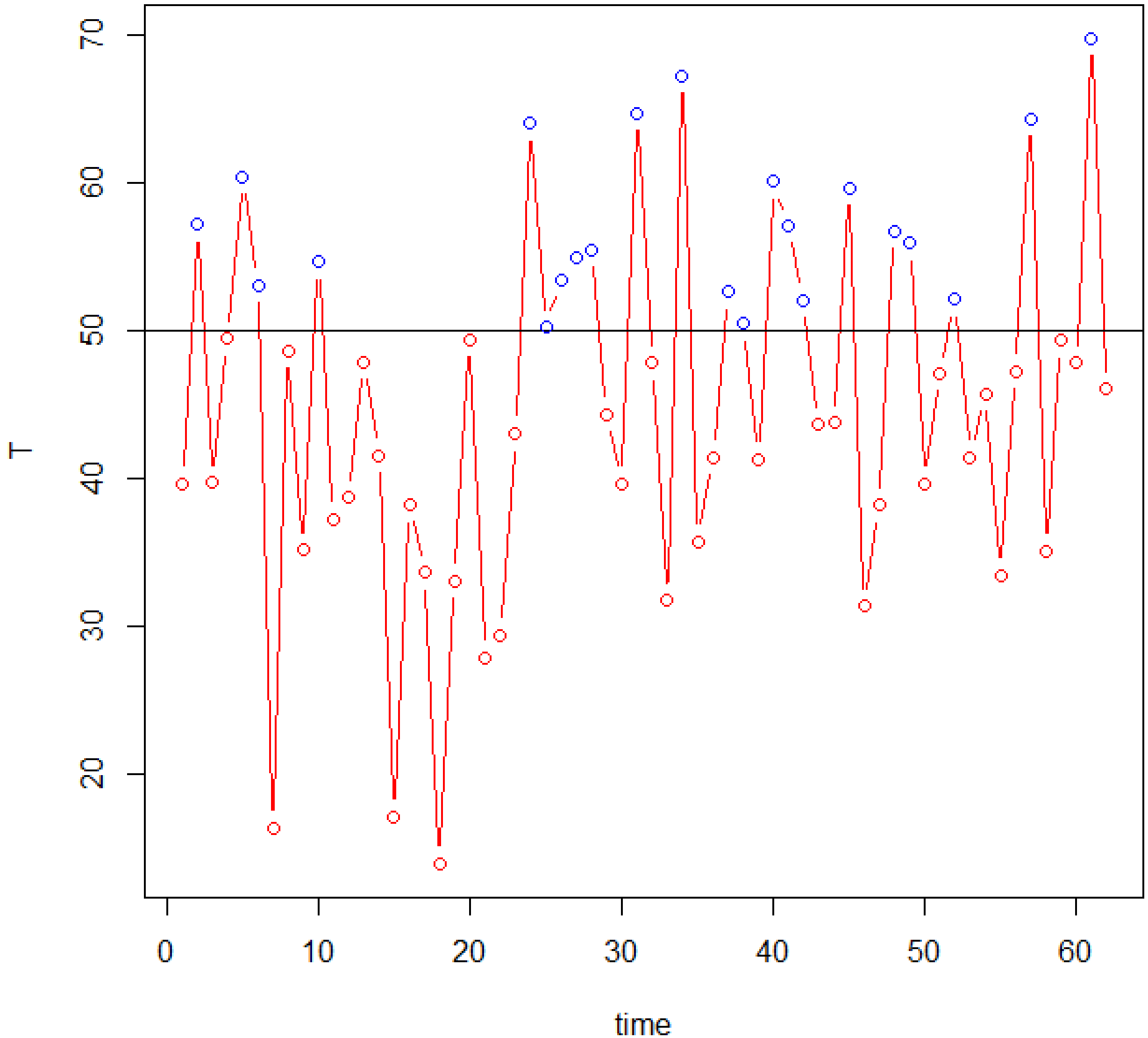
Dependence among observation

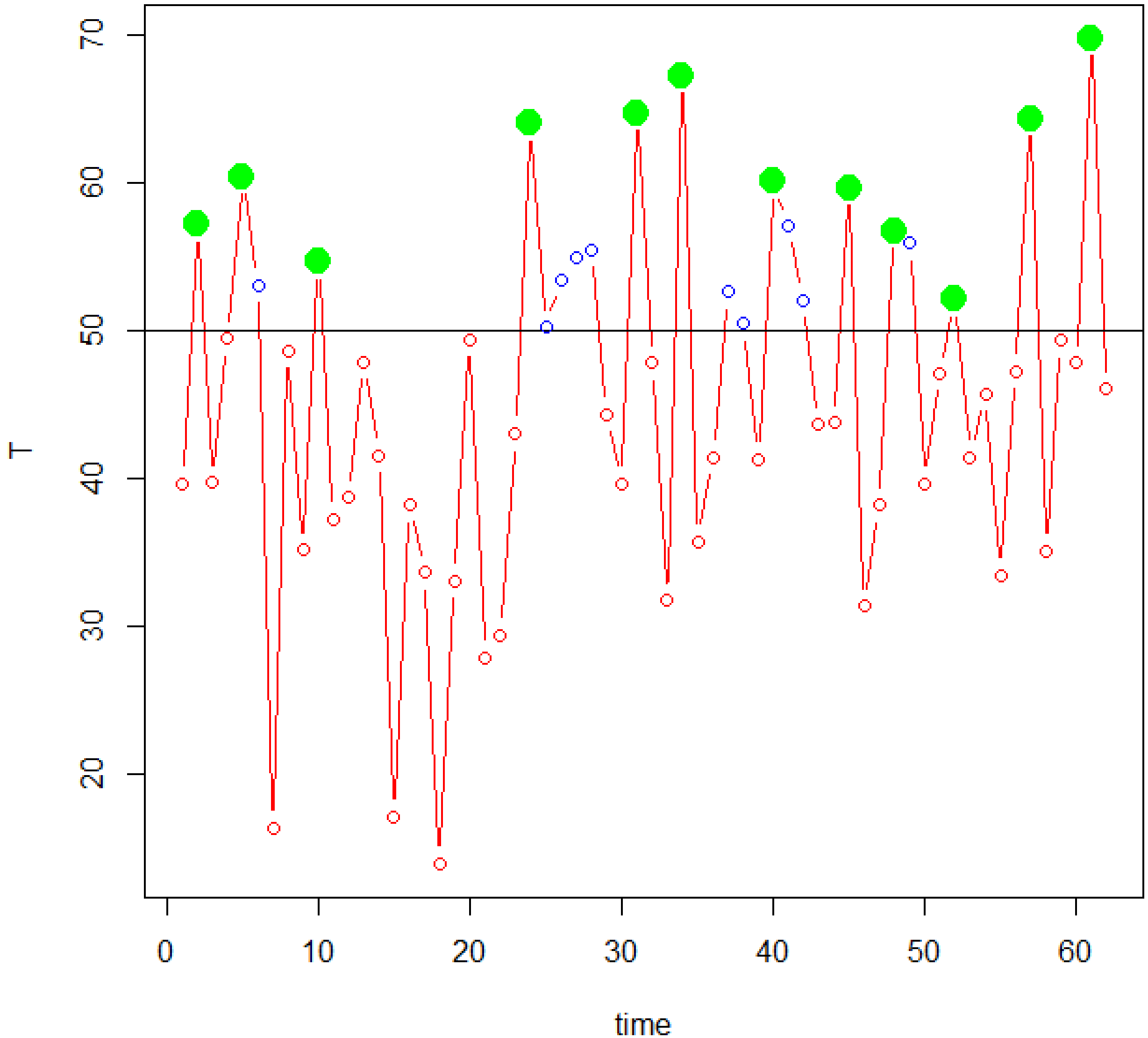
The EVT concepts introduced build on the assumption of independent identically distributed variables. Though we know that in practice most extreme values arise from a series of dependent observations.

Fortunately for sets of observations without long-range dependence the same limiting results apply for maxima of dependent time series as for independent time series (see Leadbetter et al., 1983 and Leadbetter and Rotzen 1988, for details).

In practice if a series of observations $X_1,\ldots,X_k$, with no long-range dependence of extremes has short-range dependence, which leads to extremes occurring in clusters with mean size $1/\theta$, where $0 \leq \theta \leq 1$; θ is called the **extremal index** of the process.

Dependence and Declustering




70
60
50
40
30
20
T
0
10
20
30
40
50
60
time

So if we consider a series of independent variables $X^*_1, \dots, X^*_k$ with the same marginal distribution as X_j, then $M_k = a_k[\max(X_1, \dots, X_k) - b_k]$ has a nondegenerate limiting distribution $H(y)$ if and only if $M^*_k = a_k[\max(X^*_1, \dots, X^*_k) - b_k]$ has a nondegenerate distribution $H^*(y)$, and $H(y) = [H^*(y)]^{\theta}$.

Thus the limiting distribution is unaffected by short term dependence as although the location and scale parameters differ, $H(y)$ and $H^*(y)$ have the same shape parameter .

Thus in practice the solution to clustering is to (i) identify clusters, and (ii) fit the point process model to cluster maxima.

<u>RECAP GEV:</u> As the GEV is automatically fitted to block maxima, parameters are automatically adjusted for any temporal clustering.

Stationarity vs. non-stationarity

In statistics, a **stationary process** is a stochastic process whose joint probability distribution does not change when shifted in time.

The approaches and examples discussed so far have all assumed stationarity in the underlying time series. Non-stationarity can be introduced in EVT models by expressing one or multiple parameters as a function of a covariate (e.g. time) .

Non-stationary Block maxima

As candidate model for the non-stationary GEV we can assume a model where linear trends in the location and log-transformed scale parameter [to constrain $\sigma(t) > 0$] are considered while no trend is considered in the shape parameter.

$$\mu(t) = \mu_0 + \mu_1 t,\ \ln \sigma(t) = \sigma_0 + \sigma_1 t,\ \xi(t) = \xi$$

The parameter μ_1 can be interpreted as slope of a linear trend in the center of the distribution, and the transformed parameter exp (σ_1) as the appropriate rate of change in the scale (or size of the distribution).

Such trend can be readily interpreted in terms of the corresponding time varying quantile (or 'effective' return level) which would reduce to a conventional return level (with return period $1/m$) if it would not vary with time.

If the location and/or scale parameter have linear time trends, then the effective return level would also change linearly.

So we can fit three different forms of GEV distributions to our data.

(1) A stationary GEV in which none of the parameters depends on time
(2) A nonstationary GEV in which either μ or $\ln(\sigma)$ depend on time
(3) A nonstationary GEV in which both μ and $\ln(\sigma)$ depend on time

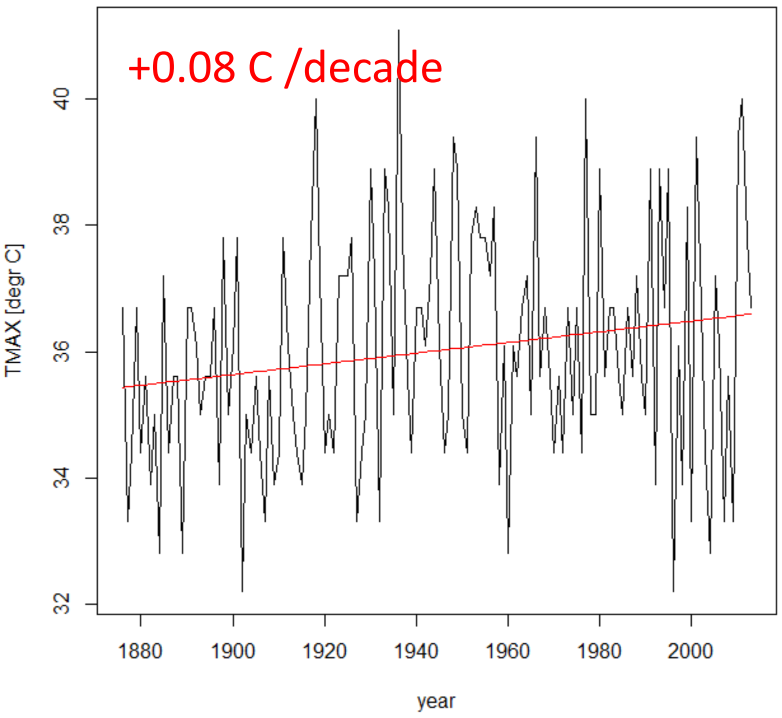
Model comparison and model selection involves the minimized negative log likelihood for the candidate models via AIC.

Non-Stationary Block Maxima - Application

Lets look on an example for our Central Park Tmax data.

We compare two models:

(1) A stationary GEV in which none of the parameters depends on time
(2) A nonstationary GEV in which μ depends on time



Non-Stationary Block Maxima - Application

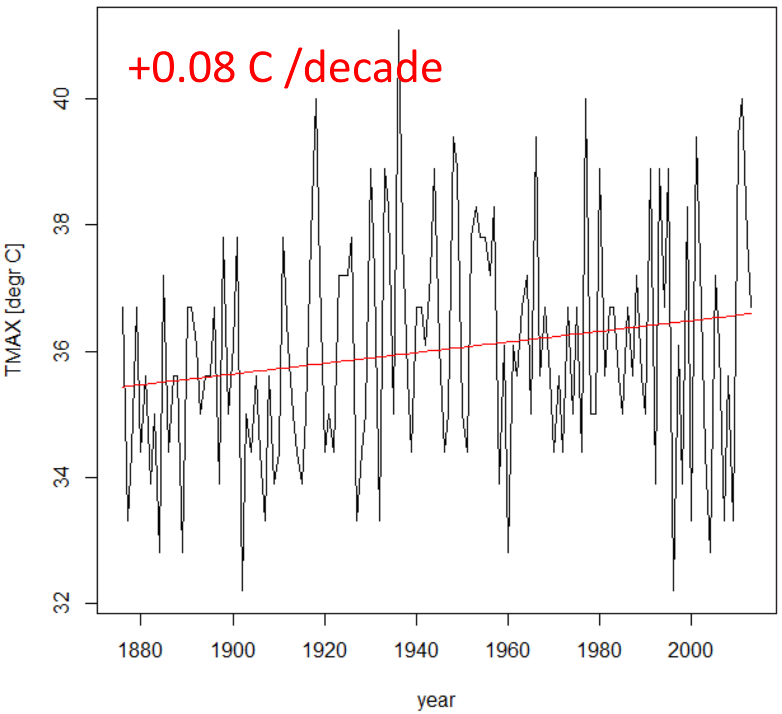
Lets look on an example for our Central Park Tmax data.

We compare two models:

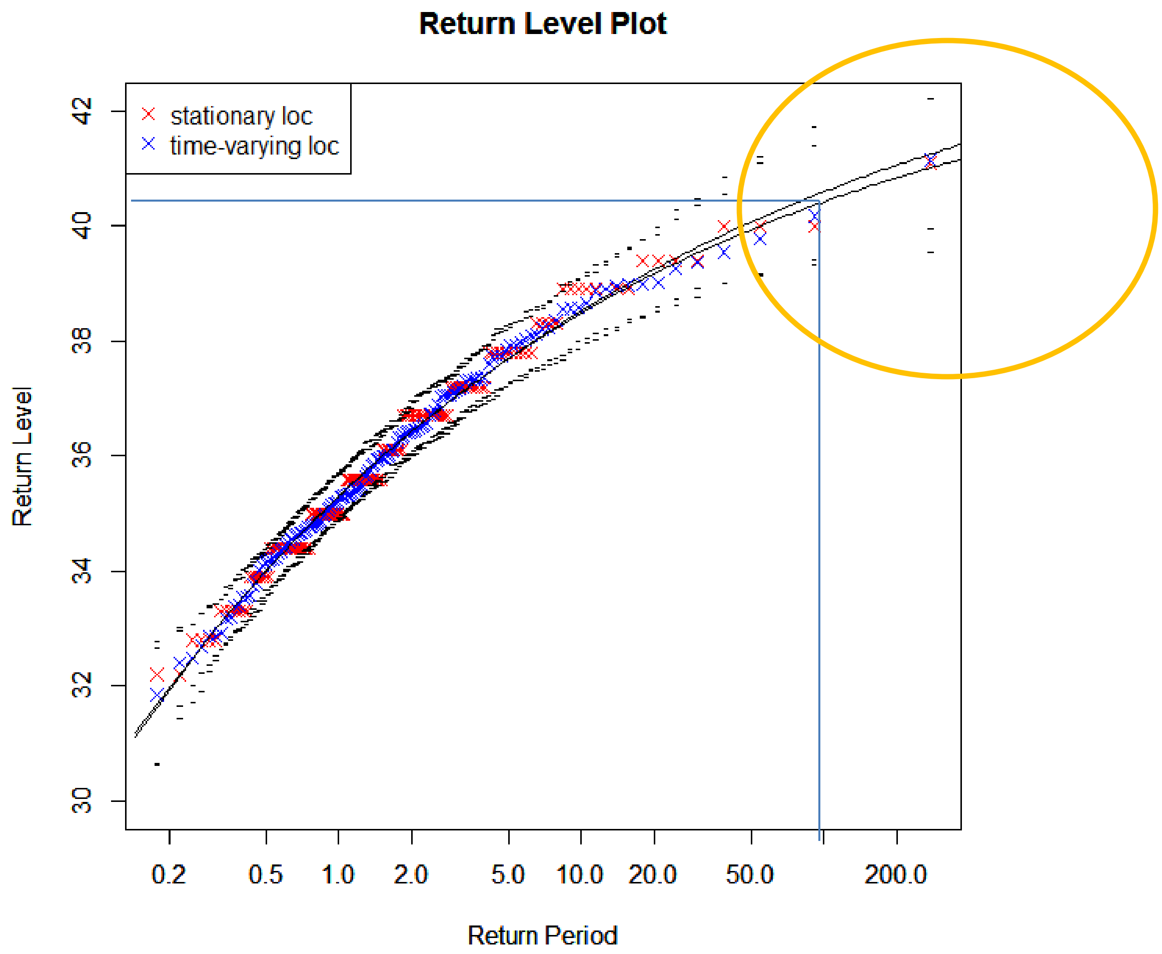
(1) A stationary GEV in which none of the parameters depends on time

(2) A nonstationary GEV in which μ depends on time

Parameter	Model 1	Model 2
μ	35.28 (±0.17)	35.31 (±0.16) + $\mu(t)$
σ	1.75 (±0.12)	1.75 (±0.12)
ξ	-0.19 (± 0.06)	-0.21 (± 0.05)
AIC	**565.73**	**564.63**
5-year RL	37.5	37.5
100-year RL	40.6	40.4



Non-Stationary Block Maxima - Application



Nonstationarity in r-largest order extremes

Introducing time dependency in parameter estimates works for r-largest order models as for block maxima. Thus as for block maxima we can consider fits of three different forms of GEV distributions to our data.

(1) A stationary GEV in which none of the parameters depends on time
(2) A nonstationary GEV in which either μ or $\ln(\sigma)$ depend on time
(3) A nonstationary GEV in which both μ and $\ln(\sigma)$ depend on time

Let's consider this in an example for the 3 warmest summer days from the Central Park record.

Non-Stationary r-largest order extremes

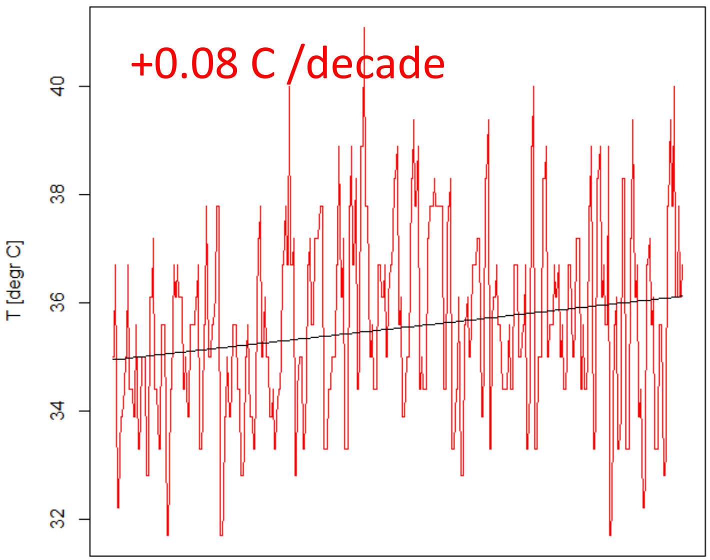
As for the block maxima approach we want to compare two models:

(1) A stationary r-largest order model where none of the GEV parameters depends on time

$$\mu(t) = \mu,\ \ln\sigma(t) = \sigma,\ \xi(t) = \xi$$

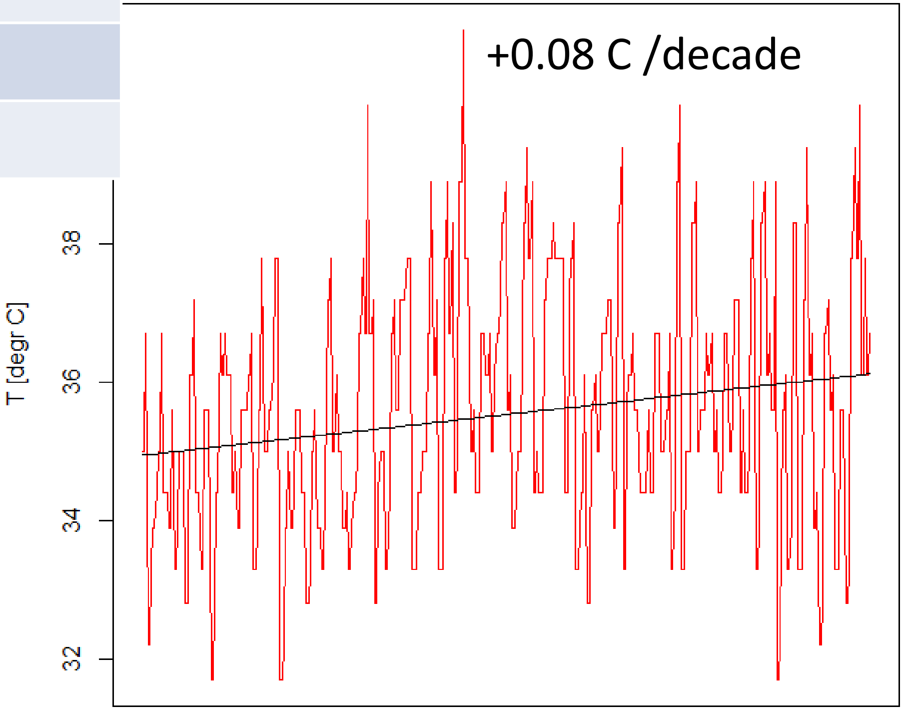
(2) A nonstationary r-largest order model where μ depends on time

$$\mu(t) = \mu_0 + \mu_1 t, \qquad \ln\sigma(t) = \sigma \quad \xi(t) = \xi$$

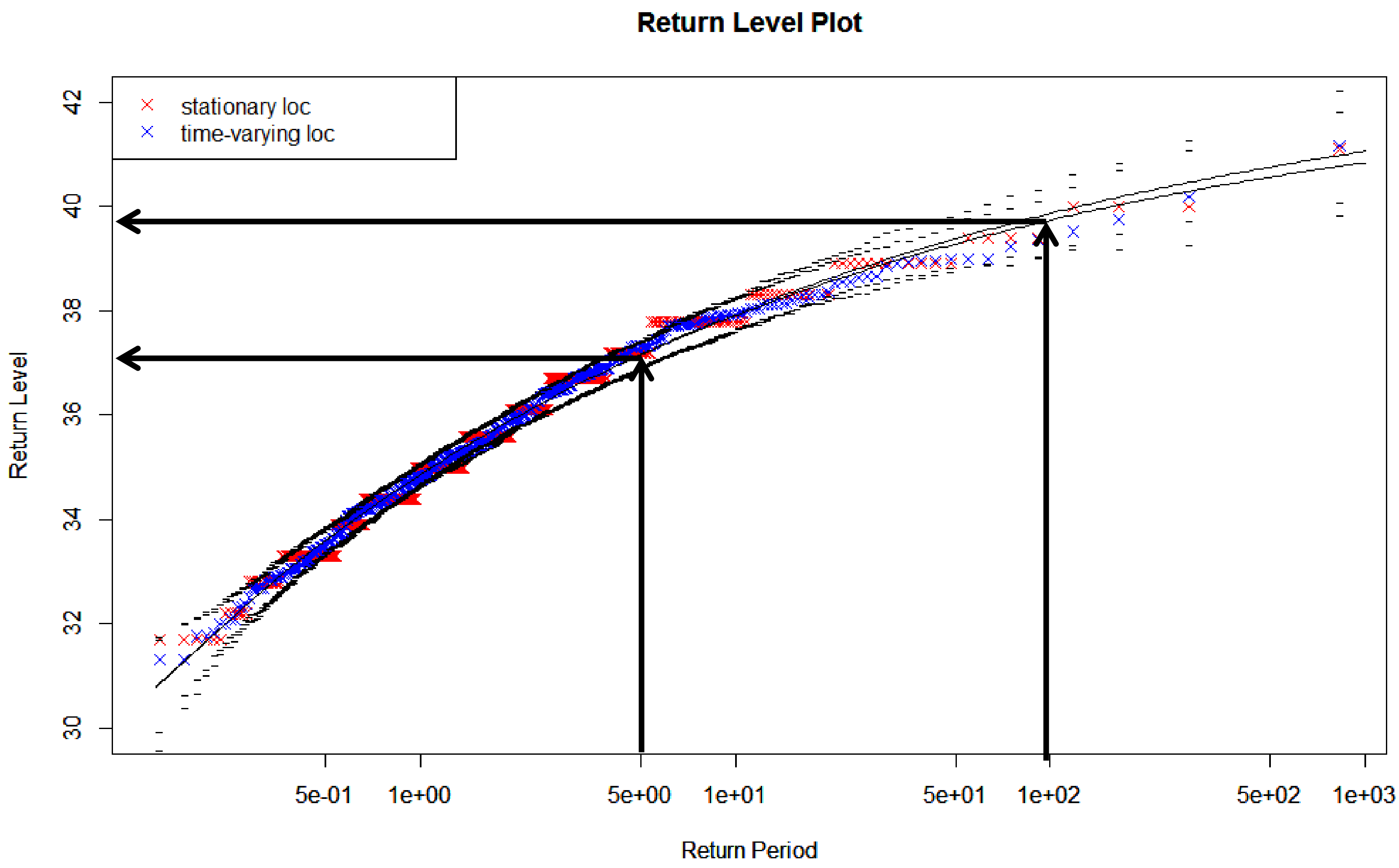


Non-Stationary r-largest order extremes - Application

Parameter	Model 1	Model 2
μ	38.84 (±0.09)	38.86 (±0.16) + μ(t)
σ	1.69 (±0.06)	1.69 (±0.06)
ξ	-0.21 (± 0.03)	-0.22 (± 0.03)
AIC	**1653.7**	**1644.1**
5-year RL	37.0 (37.5)	37.0 (37.5)
100-year RL	39.8 (40.6)	39.7 (40.4)



Non-Stationary r-largest order extremes - Application



Non-stationarity in Peak over Threshold models

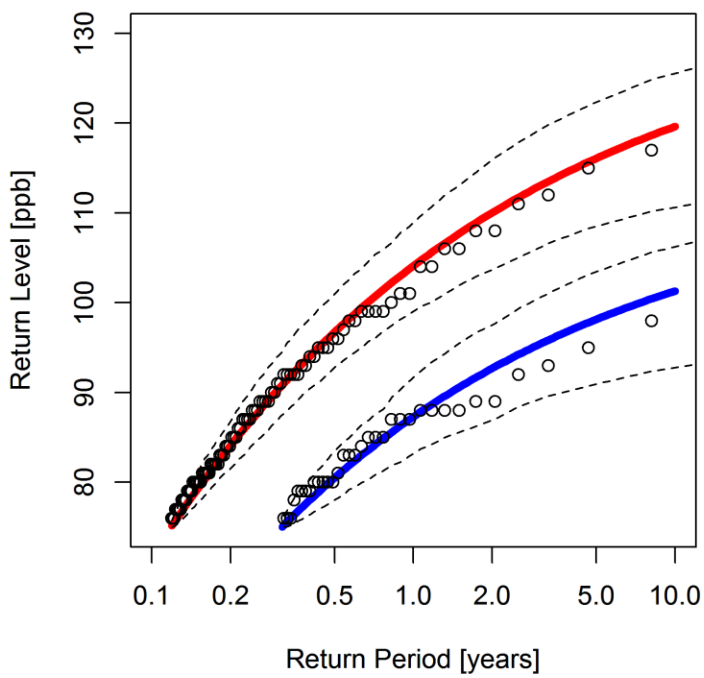
Frequently we are interested in extremes defined as exceedances of a certain threshold and we know that the POT model is the suitable EVT model for such type of analysis. Non-stationarity can be addressed in POT models though a bit caution is needed as:

(1) Normally the threshold we are interested in is a fixed quantity, nevertheless a time-varying threshold can be introduced in POT models.
(2) The scale parameter is dependent on u thus care must be given in the interpretation of σ when using a time varying threshold.
(3) If we are only interested in the change in return levels due to changes in the evolution of the data (e.g., a trend) we can assess changes by fitting the GPD to different time periods an compare the RL estimates.

Let's look on mda8 O_3 return periods for two different time periods

(1) 1988-2000

(2) 2001-2013



Communicating risk in a non-stationary setting – notes on return level estimates under non-stationarity

Let $F(y)$ be the distribution function of $M(y)$. In any particular year there is still a one to one relationship between a probability of exceedance and a high quantile.

So for a given level of interest r, it is straightforward to express the yearly risk in terms of probability.

$$p(y) = P\left(M_y > r\right) = 1 - F_y(r).$$

Once F_y is estimated it is straightforward to provide yearly estimates of the probability of an exceedance $p(y)$.

Risk calculations however proceed in the opposite direction. Normally one starts with a return period in the stationary case and finds the corresponding level.

Inverting the previous procedure arriving at a probability of exceedance p, one starts with the probability of exceedance p and solves

$$F_y\left(r_p(y)\right) = 1 - p \; .$$

The exceedance level $r_p(y)$ changes with every year thus communicates clearly the changing nature of risk.

Return level estimates under non-stationarity

If we define the return level r_m as the expected waiting time until an exceedance occurs in m years, then r_m is the solution to equation

$$m = 1 + \sum_{i=1}^{\infty} \prod_{y=1}^{i} F(y)(r_m)$$

It shall be noted that as it cannot be written as geometric series, solving for r_m is not straightforward. For the case $F_y(r)$ is monotonically decreasing as $y \to \infty$ (extremes are getting more extreme) it is possible to bound the right hand side of the equation above (see Cooley 2013 for more details).

Return level estimates under non-stationarity

The other interpretation of an m-year return period under the stationary case was that the expected number of exceedances in m years is one.

To extend this for the non-stationary case we aim to find the level r_m for which the expected number of exceedances in m years is one.

So if we let N be the number of exceedances that occur in the m years beginning with year $y = 1$ and ending at year $y = m$. Then as the probability of an exceedance is no longer constant from year to year and we define the m-year return level r_m to be the solution to the equation

$$1 = \sum_{y=1}^{m} \left(1 - F_y(r_m)\right).$$

Box 2.4, Table 1 | Definitions of extreme temperature and precipitation indices used in IPCC (after Zhang et al., 2011). The most common units are shown but these may be shown as normalized or relative depending on application in different chapters.

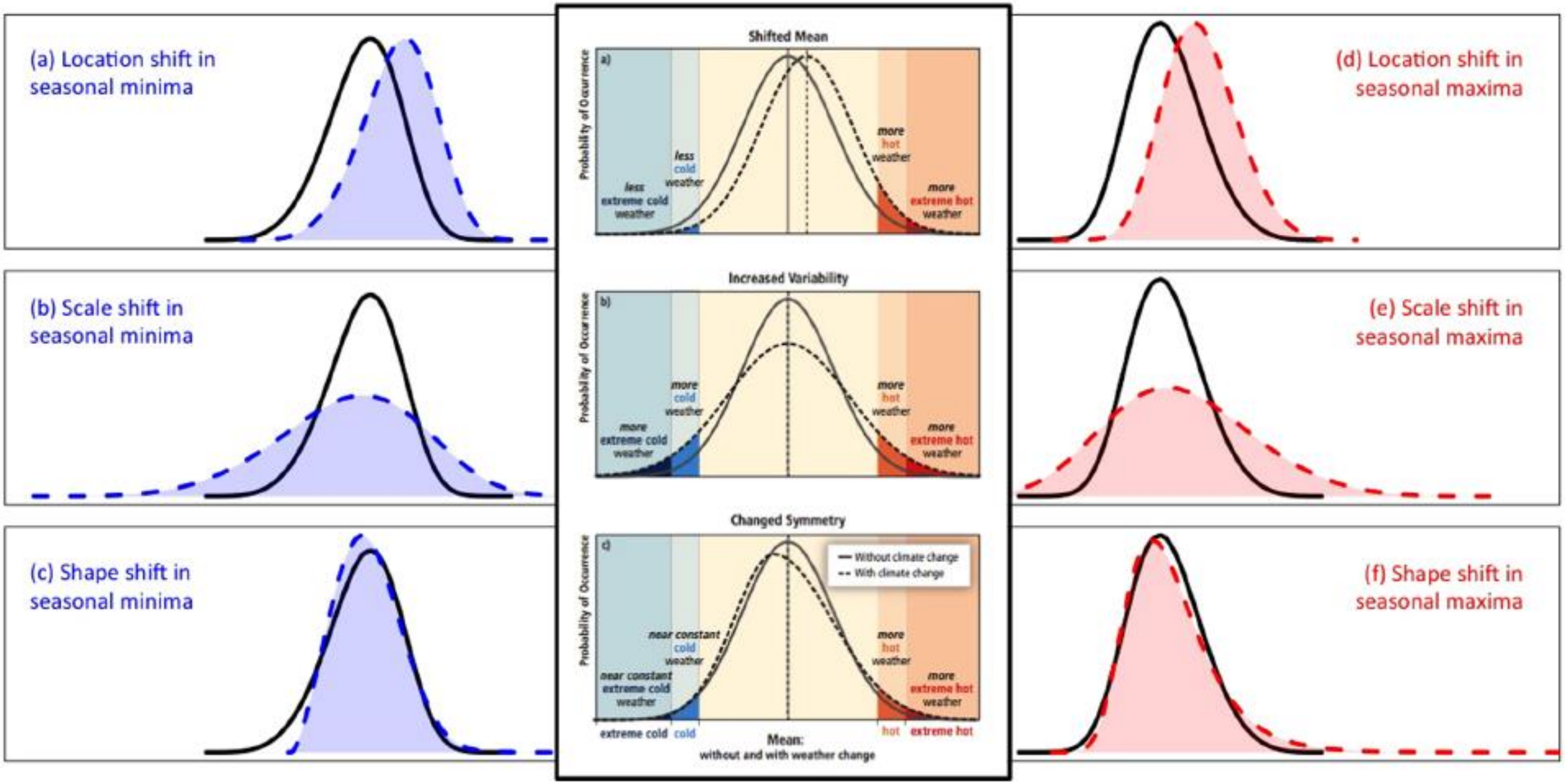
Index	Descriptive name	Definition	Units	Figures/Tables	Section
TXx	Warmest daily Tmax	Seasonal/annual maximum value of daily maximum temperature	°C	Box 2.4, Figure 1, Figures 9.37, 10.17, 12.13	Box 2.4, 9.5.4.1, 10.6.1.1, 12.4.3.3
TNx	Warmest daily Tmin	Seasonal/annual maximum value of daily minimum temperature	°C	Figures 9.37, 10.17	9.5.4.1, 10.6.1.1
TXn	Coldest daily Tmax	Seasonal/annual minimum value of daily maximum temperature	°C	Figures 9.37, 10.17, 12.13	9.5.4.1, 10.6.1.1, 12.4.3.3
TNn	Coldest daily Tmin	Seasonal/annual minimum value of daily minimum temperature	°C	Figures 9.37, 10.17, 12.13	9.5.4.1, 10.6.1.1
TN10p	Cold nights	Days (or fraction of time) when daily minimum temperature <10th percentile	Days (%)	Figures 2.32, 9.37, 10.17 Tables 2.11, 2.12	2.6.1, 9.5.4.1, 10.6.1.1, 11.3.2.5.1
TX10p	Cold days	Days (or fraction of time) when daily maximum temperature <10th percentile	Days (%)	Figures 2.32, 9.37, 10.17, 11.17	2.6.1, 9.5.4.1, 10.6.1.1, 11.3.2.5.1,
TN90p	Warm nights	Days (or fraction of time) when daily minimum temperature >90th percentile	Days (%)	Figures 2.32, 9.37, 10.17 Tables 2.11, 2.12	2.6.1, 9.5.4.1, 10.6.1.1, 11.3.2.5.1
TX90p	Warm days	Days (or fraction of time) when daily maximum temperature >90th percentile	Days (%)	Figures 2.32, 9.37, 10.17, 11.17 Tables 2.11, 2.12	2.6.1, 9.5.4.1, 10.6.1.1, 11.3.2.5.1,
FD	Frost days	Frequency of daily minimum temperature <0°C	Days	Figures 9.37, 12.13 Table 2.12	2.6.1, 9.5.4.1, 10.6.1.1, 12.4.3.3
TR	Tropical nights	Frequency of daily minimum temperature >20°C	Days	Figures 9.37, 12.13	9.5.4.1, 12.4.3.3
RX1day	Wettest day	Maximum 1-day precipitation	mm	Figures 9.37, 10.10 Table 2.12, 12.27	2.6.2.1, 9.5.4.1, 10.6.1.2, 12.4.5.5
RX5day	Wettest consecutive five days	Maximum of consecutive 5-day precipitation	mm	Figures 9.37, 12.26, 14.1	9.5.4.1, 10.6.1.2, 12.4.5.5, 14.2.1
SDII	Simple daily intensity index	Ratio of annual total precipitation to the number of wet days (≥1 mm)	mm day^{-1}	Figures 2.33, 9.37, 14.1	2.6.2.1, 9.5.4.1, 14.2.1
R95p	Precipitation from very wet days	Amount of precipitation from days >95th percentile	mm	Figures 2.33, 9.37, 11.17 Table 2.12	2.6.2.1, 9.5.4.1, 11.3.2.5.1
CDD	Consecutive dry days	Maximum number of consecutive days when precipitation <1 mm	Days	Figures 2.33, 9.37, 12.26, 14.1	2.6.2.3, 9.5.4.1, 12.4.5.5, 14.2.1

Suitability of EVT models for climate extremes

Block Maximum Models:
Warmest daily Tmax (TXx), Warmest daily Tmin (TNx)
Coldest daily Tmax (TXn), Coldest daily Tmin (TNn)
Wettest day (RX1day)
Wettest consecutive 5 days (RX5day)

R-largest order models:
Cold days (TX10p), Cold Nights (TN10p)
Warm days (TX90p), Warm Nights (TN90p)

Peak-Over-Threshold Models:
Frost days (FD), Tropical Nights (TR)
Cold days (TX10p), Cold Nights (TN10p)
Warm days (TX90p), Warm Nights (TN90p)



Kodra & Ganguly, Nature Scientific Reports, 2014

For those who are using R

Packages for extreme value analysis

evd
ismev
POT
extRemes