IPCC workshop on extreme weather and climate events,
11-13 June 2002, Beijing.

Estimation and attribution of changes in extreme weather and climate events

Dr. David B. Stephenson
Department of Meteorology
University of Reading
www.met.rdg.ac.uk/cag

Plan of the talk

1. Introduction
2. Descriptive methods
3. Understanding the changes
4. Modelling the tail of the pdf
5. Recommendations

1. Introduction …

The definition problem:

Extreme events can be defined by:

- Maxima/minima
- Magnitude
- Rarity
- Impact/losses

Gare Montparnasse, 22 October 1895

“Man can believe the impossible, but man can never believe the improbable.” - Oscar Wilde

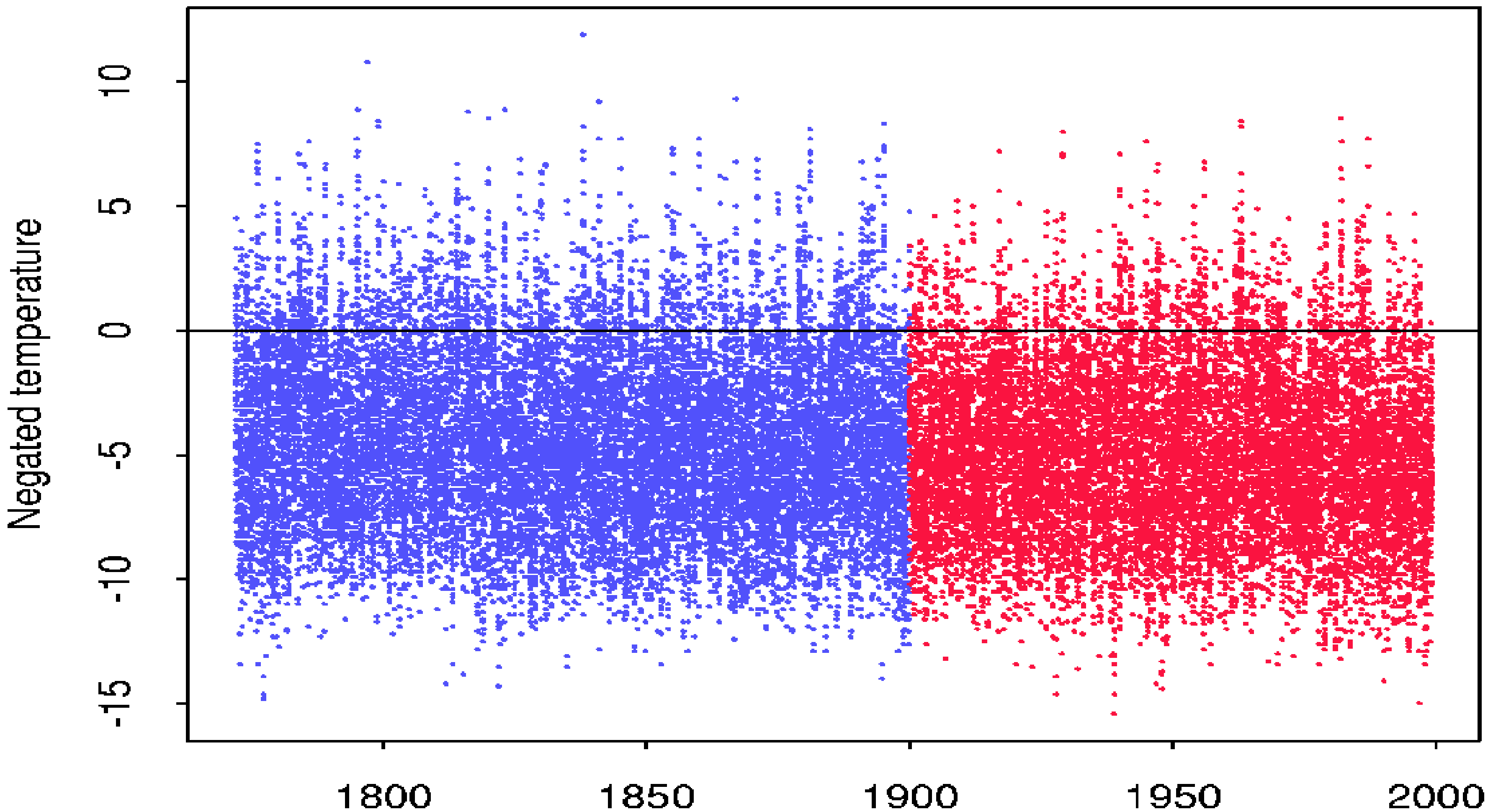
IPCC 2001 definition of extreme event:

"An extreme weather event is an event that is rare within its statistical reference distribution at a particular place. Definitions of "rare" vary, but an extreme weather event would normally be as rare or rarer than the 10th or 90th percentile."

Daily CET winter temperatures (Nov-Mar)

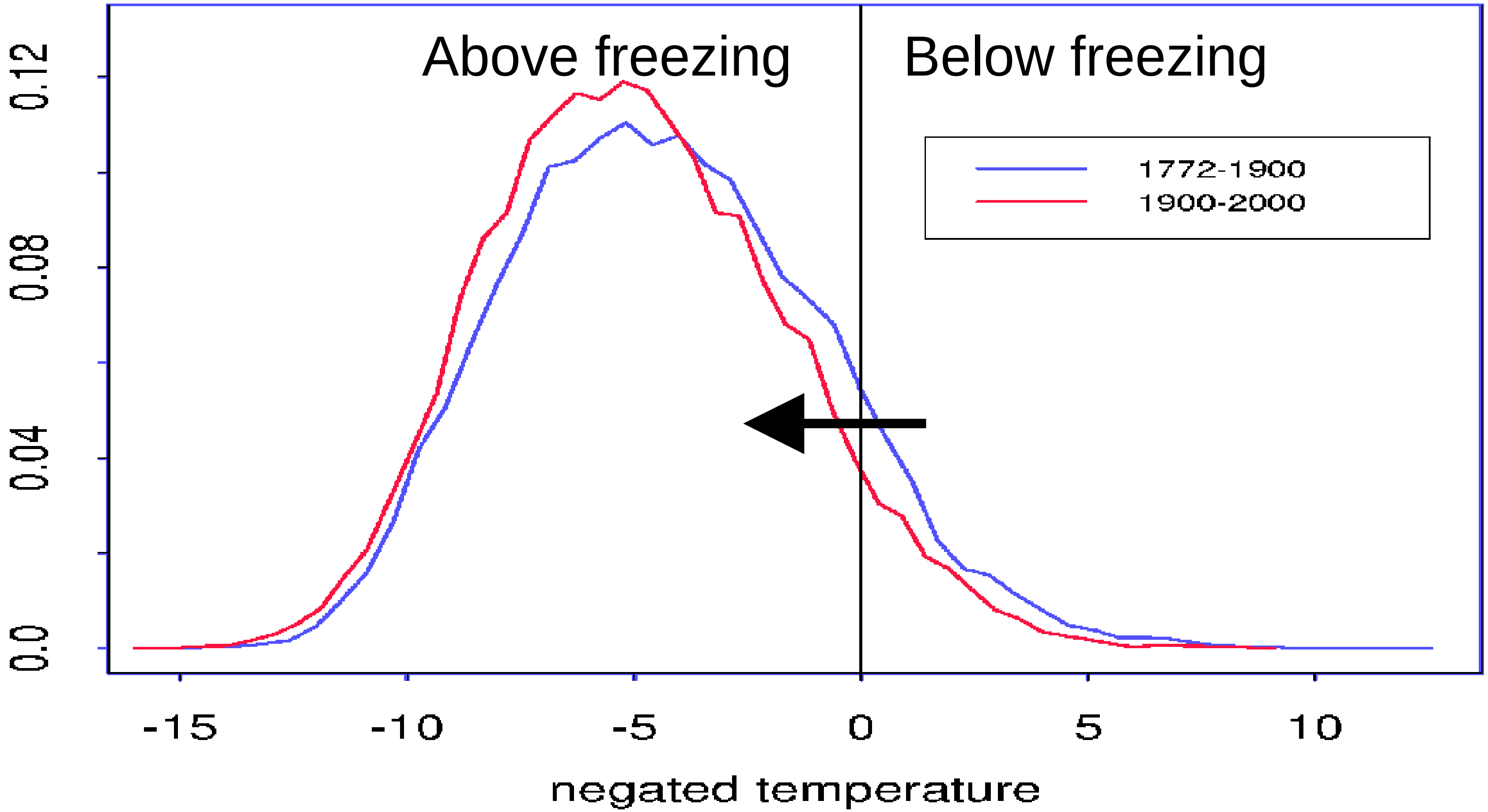
Sample 1: 1772-1900 n=19232 values

Sample 2: 1900-2000 n=14933 values



2. Descriptive methods ...

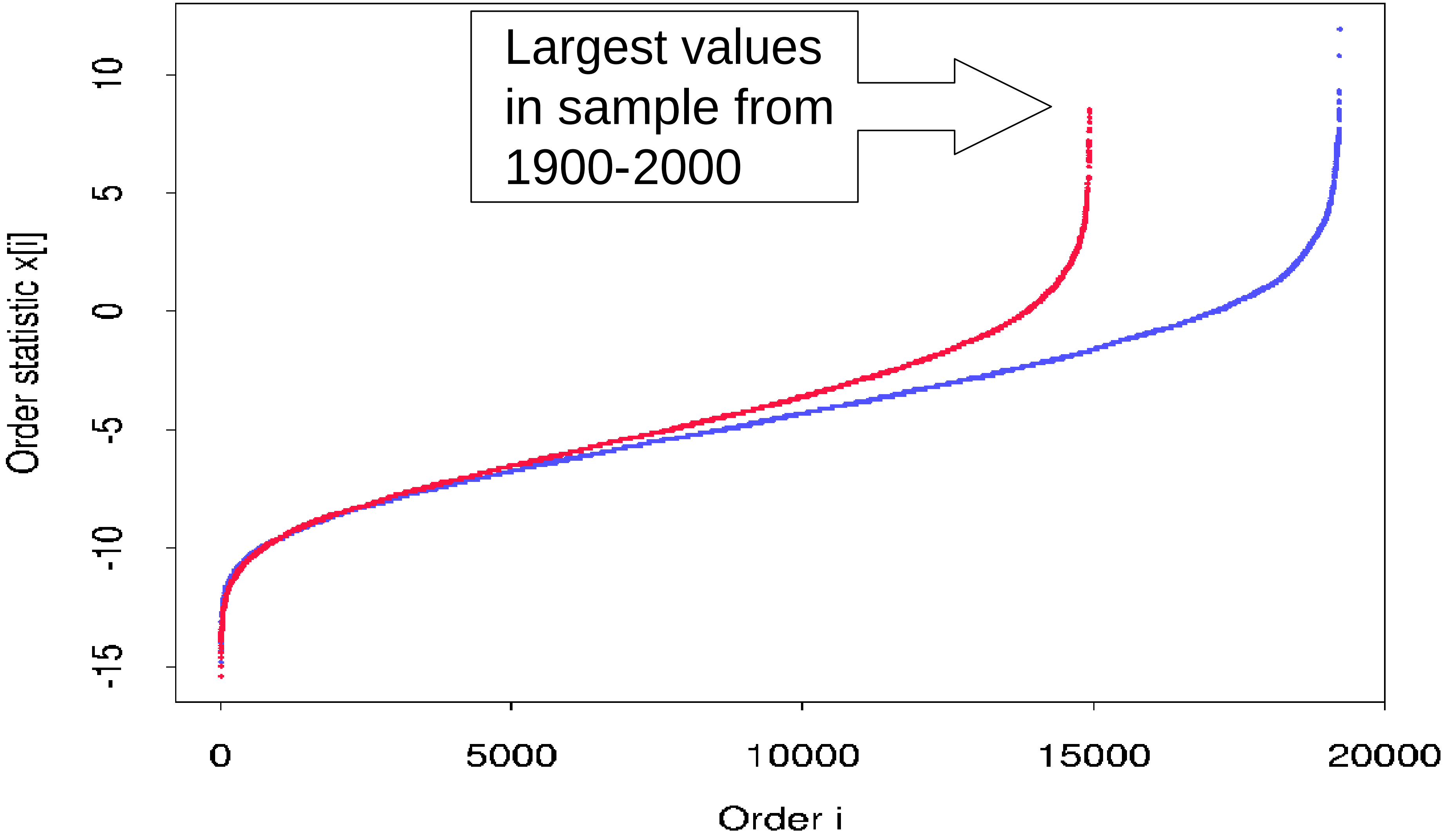
Change in the p.d.f. of winter temp.



Note shift in probability density for cold temperatures

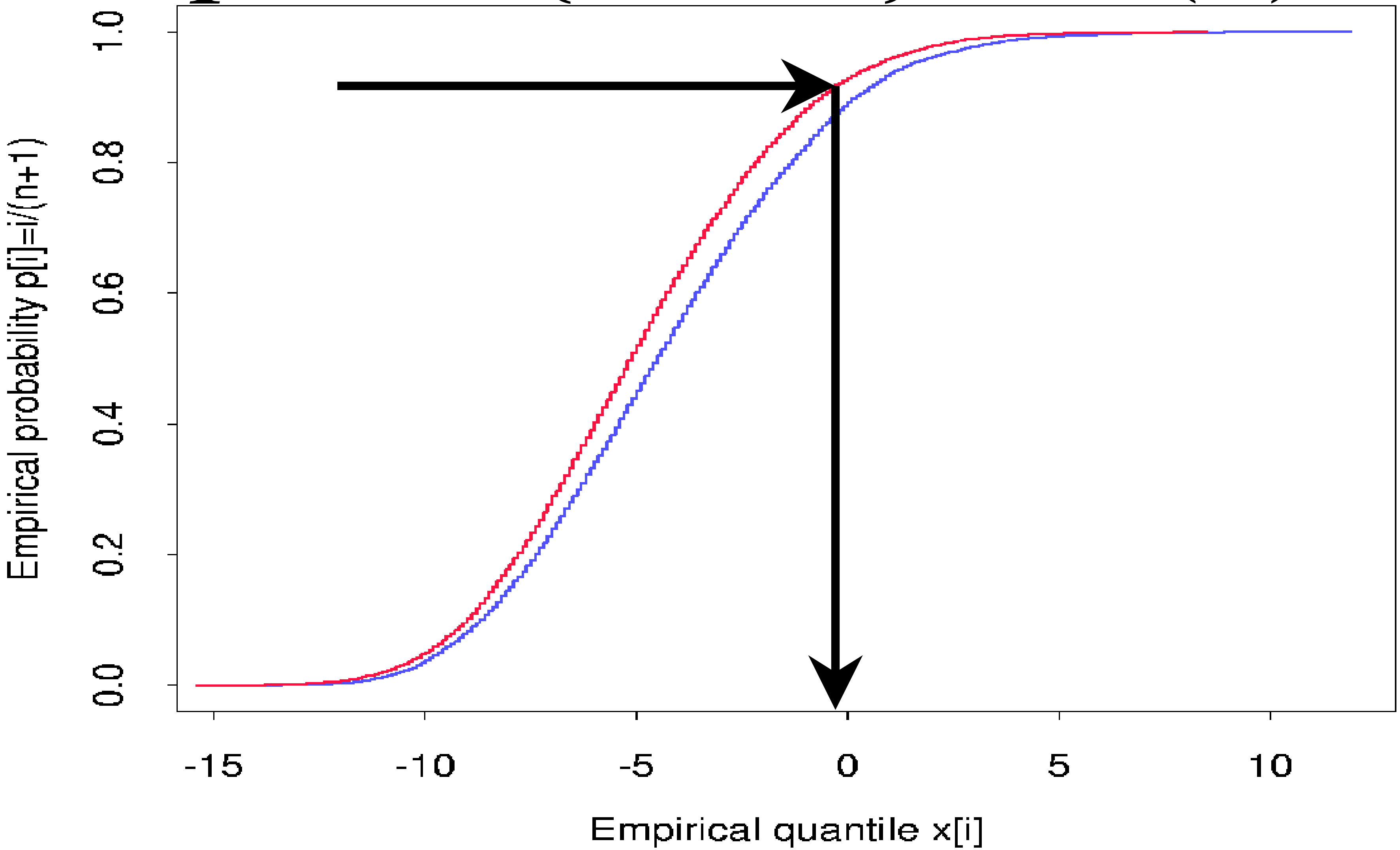
Order statistics $x_{[i]}$

Simply rank sample values in ascending order …

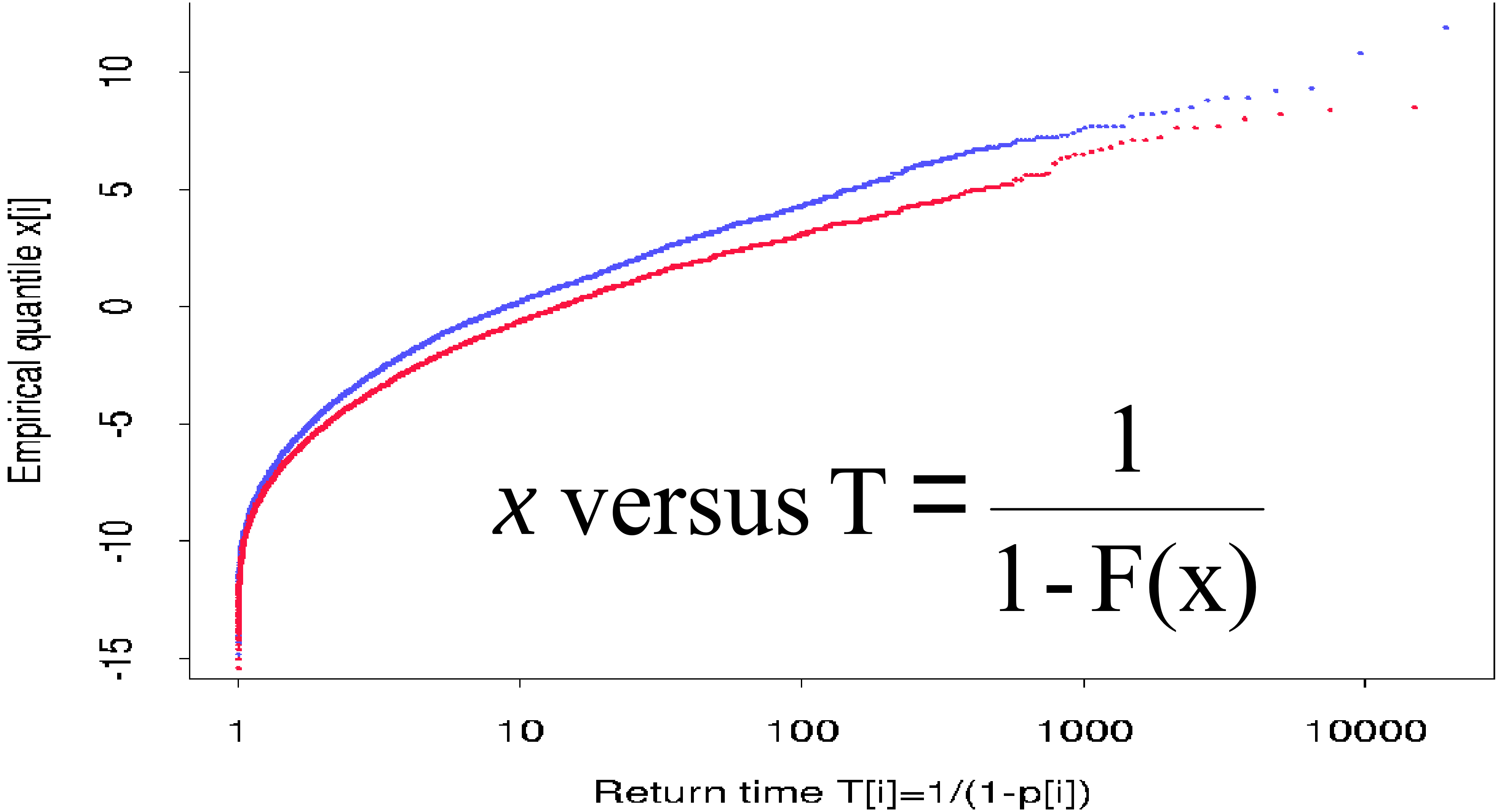


Empirical Distribution Function (e.d.f.)

$$p = \Pr\{X \leq x\} = F(x)$$

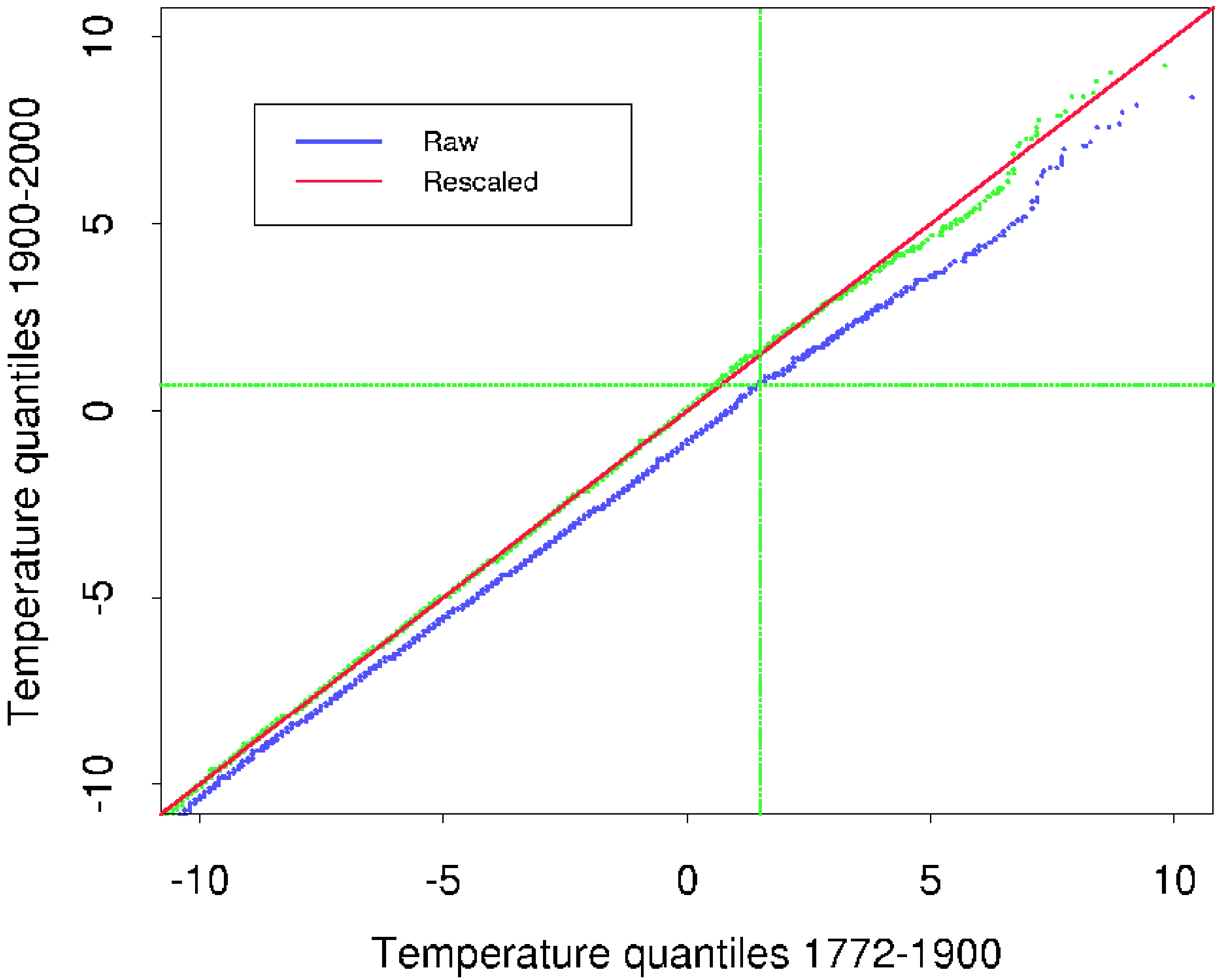


Empirical return level plot



A way of plotting the distribution function x vs. 1/(1-F)

Empirical quantile-quantile plots



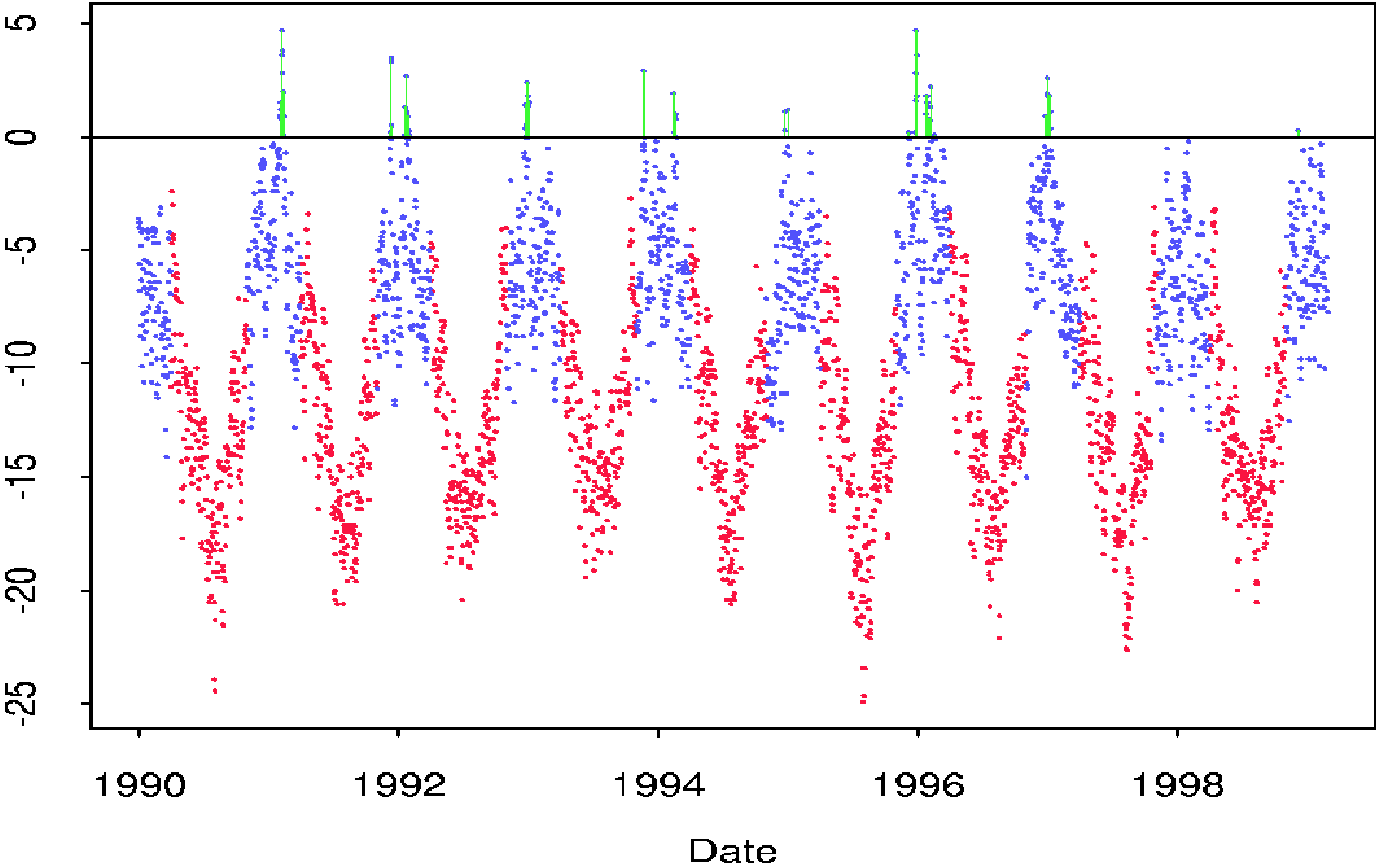
Note:

- quantiles differ
- not just a shift
  due to means

1772-1900:
- mean=-4.31 deg C
- stdev= 3.45 deg C

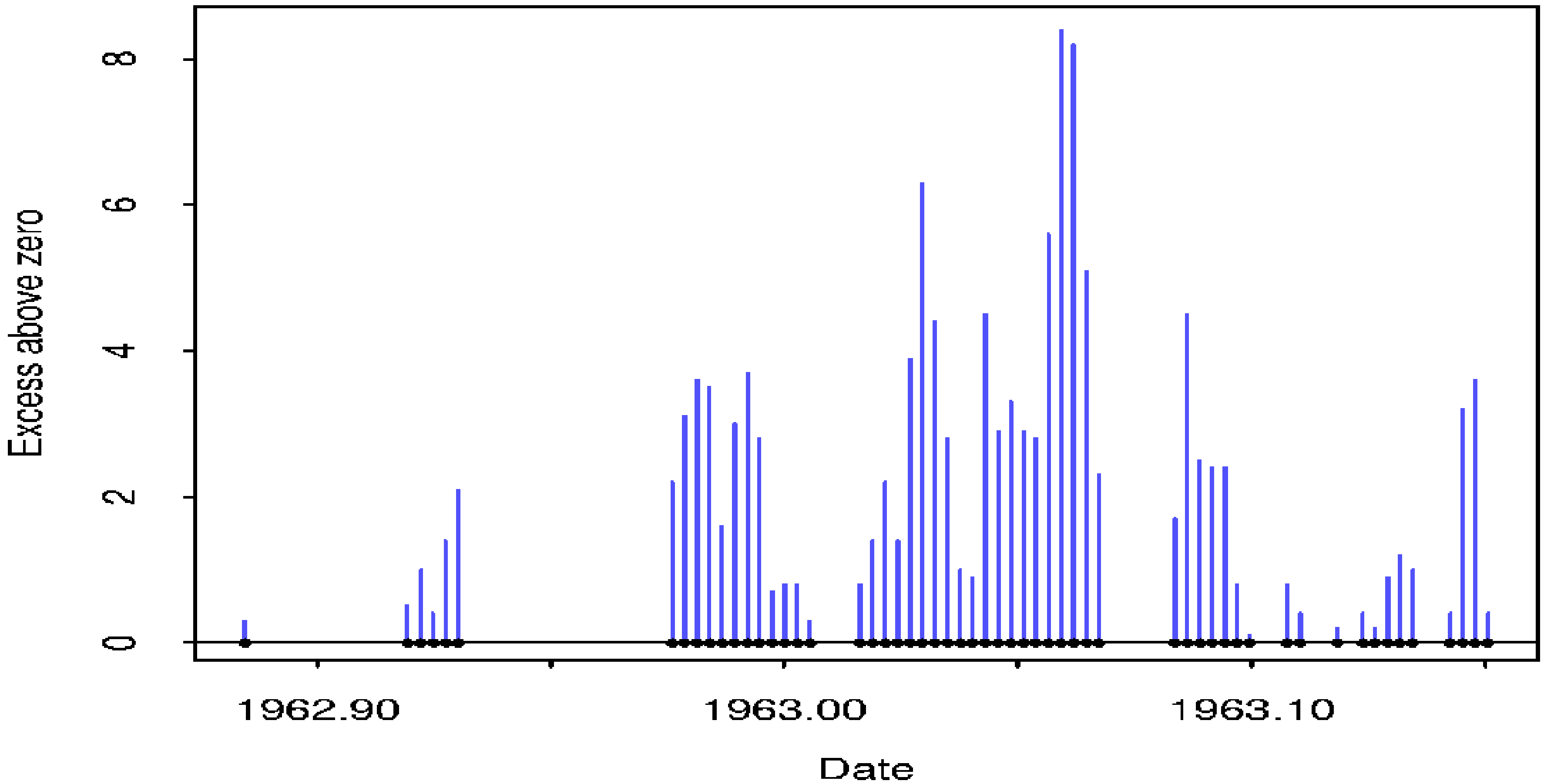
1900-2000:
- mean=-4.94 deg C
- stdev= 3.26 deg C

Negated Central England Temperature



Marked point process for exceedances

Marked point process = random point process in time
+ random amount at each event

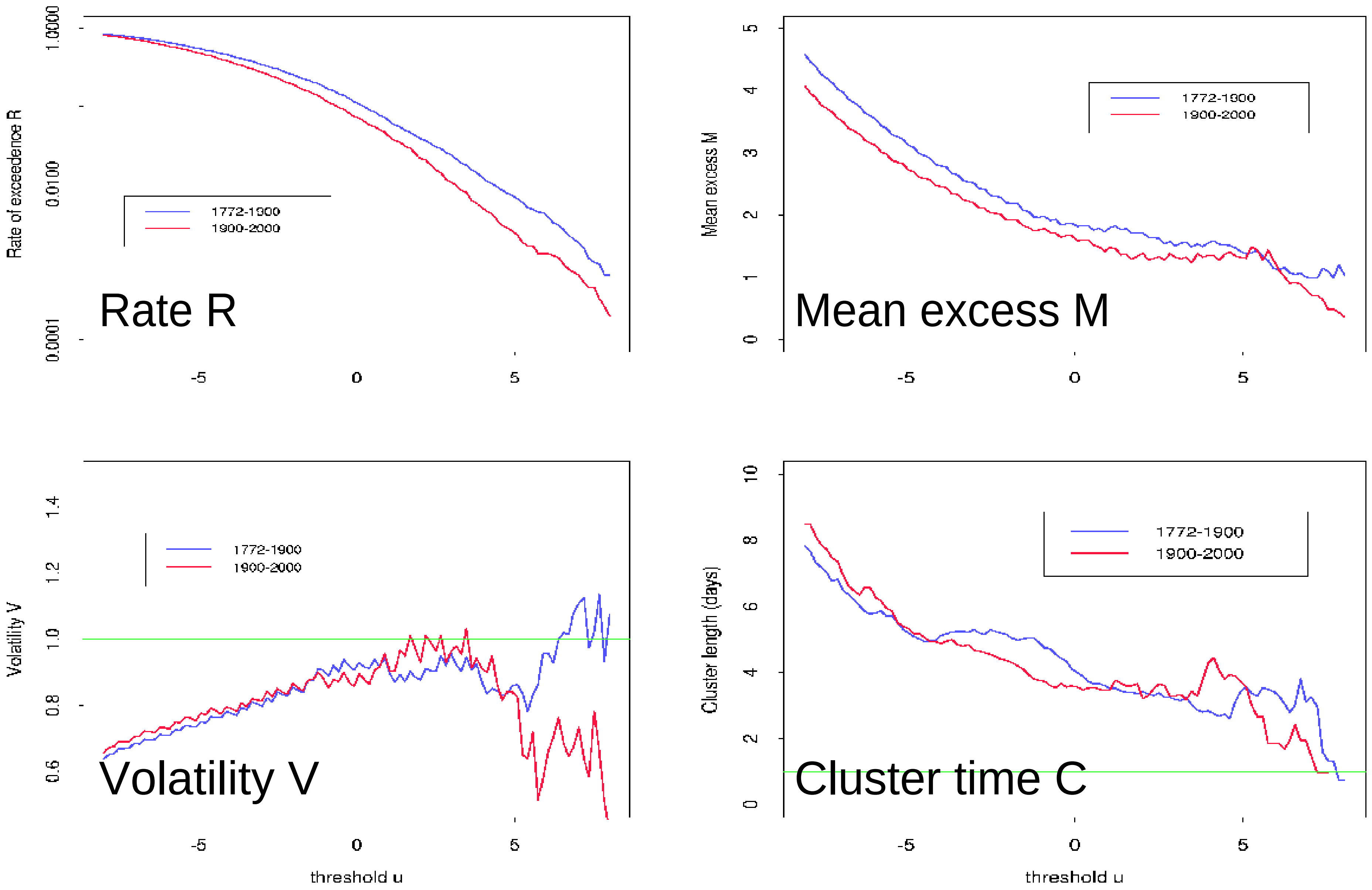


Principal attributes of exceedances

- Rate of exceedance $R = \Pr\{X > u\}$
- Mean excess $M = E(X - u \mid X > u)$
- Volatility $V = \mathrm{Stdev}(X - u \mid X > u)/M$
- Clustering in time $L = 1/\text{Extremal Index}$

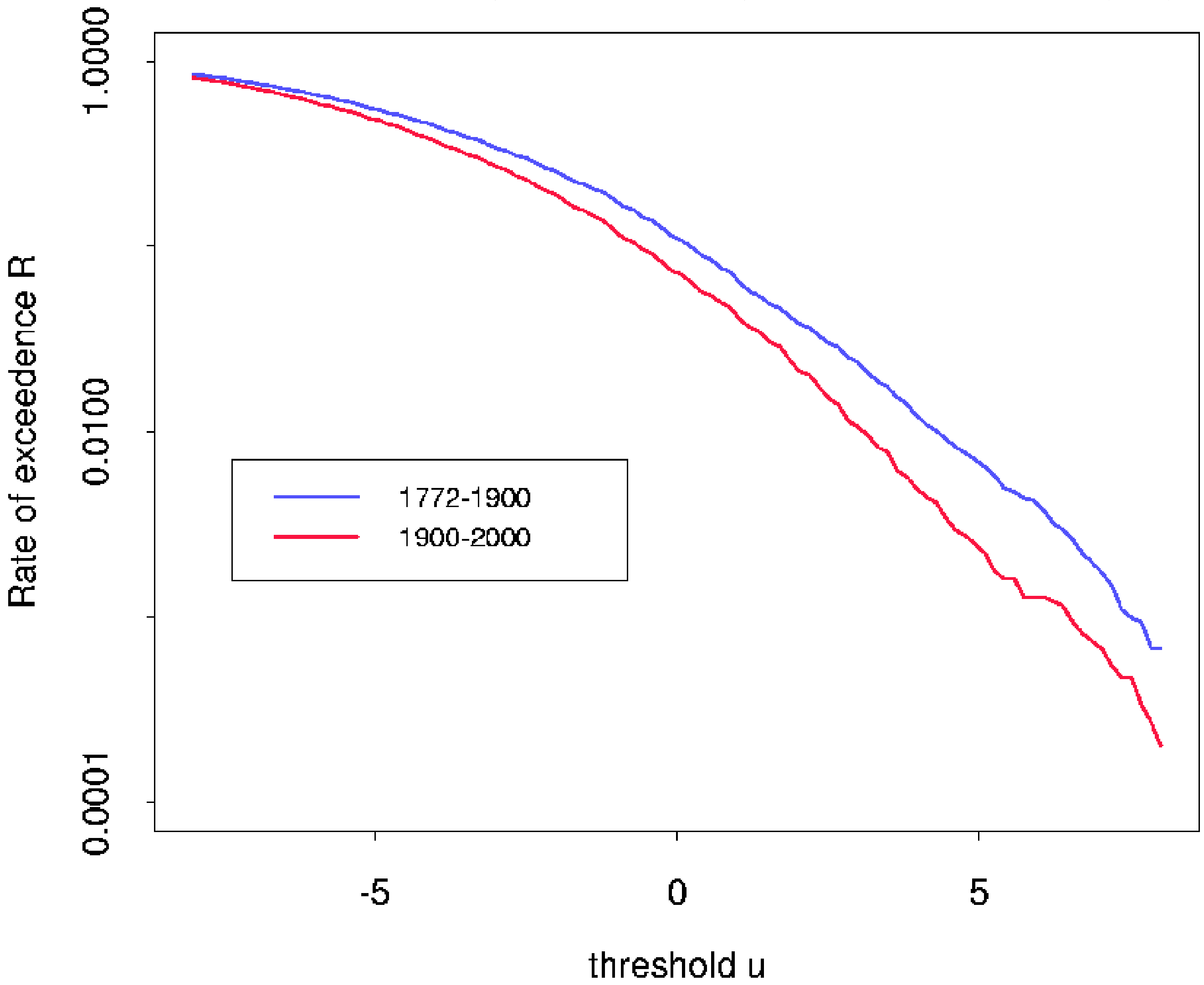
Note: In many weather/climate studies only rate is discussed yet the other attributes also merit attention!

Attributes as a function of threshold



Changes in the rate of exceedance R

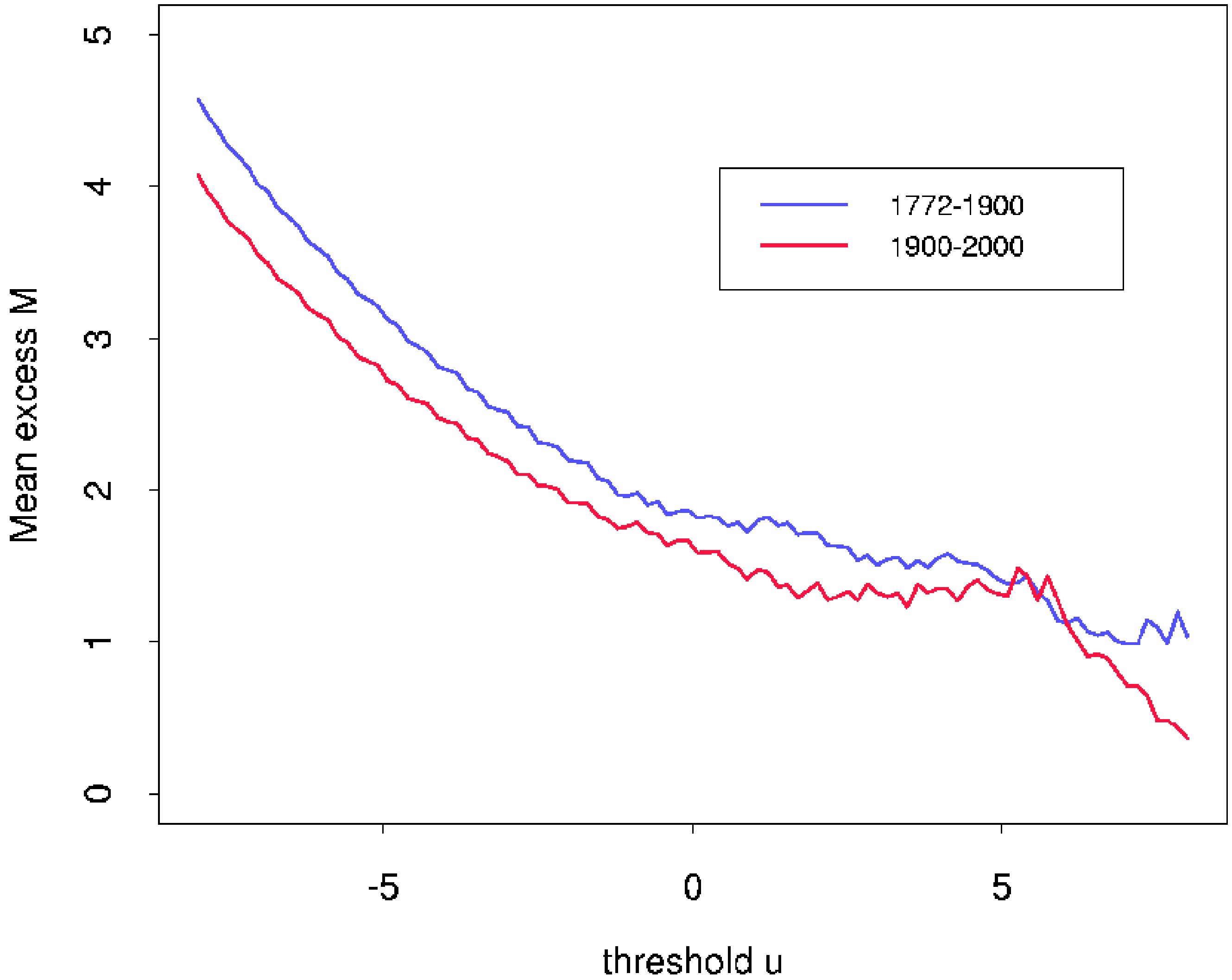
$$R = \Pr\{X > u\} = 1 - F(u)$$



Note reduced rate of exceedance in the later period.

Changes in the mean excess

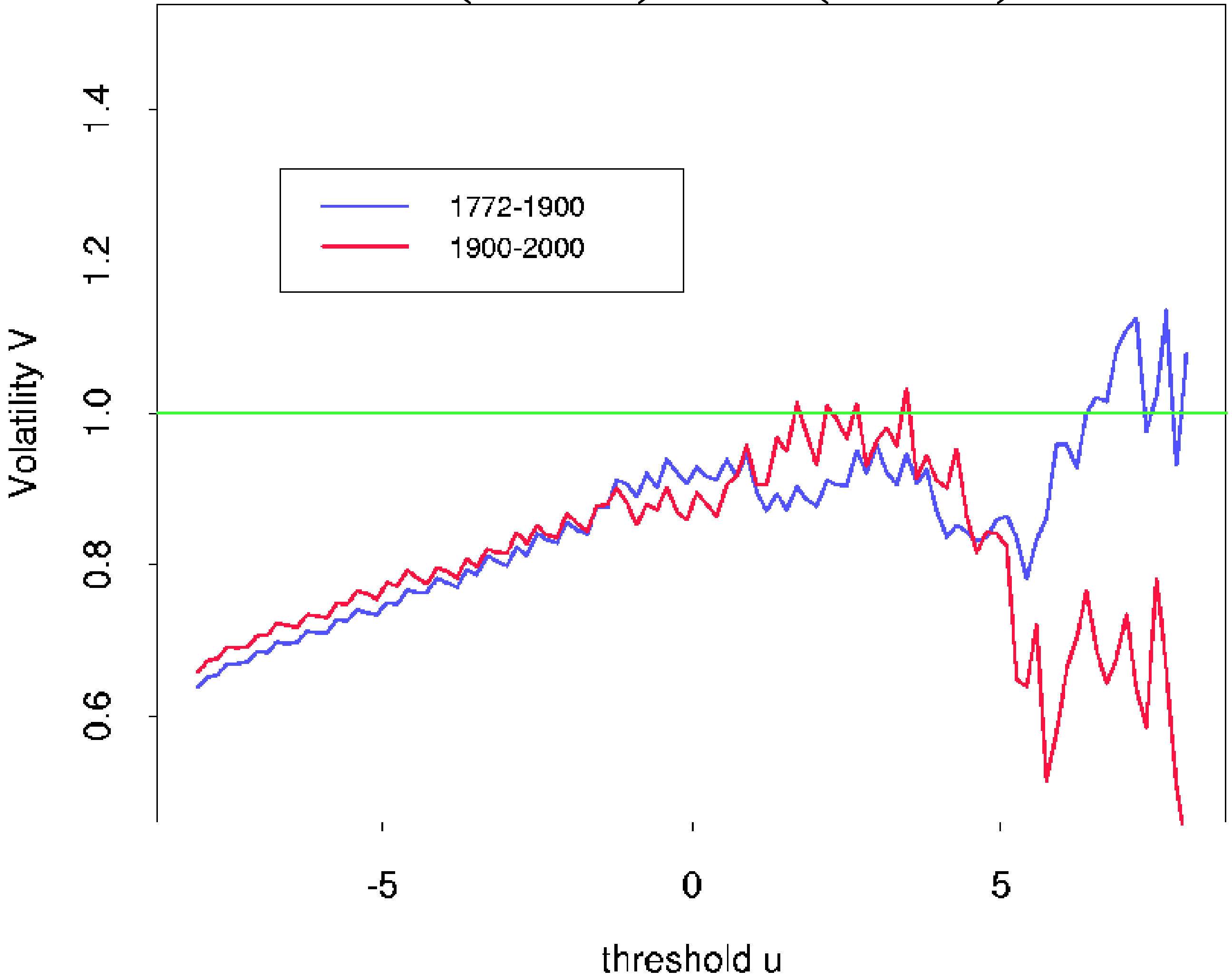
$$M = \text{mean excess} = E(X - u \mid X > u)$$



Changes in the volatility of excesses

$$\text{Volatility} = \text{Coefficient of Variation of excesses}$$

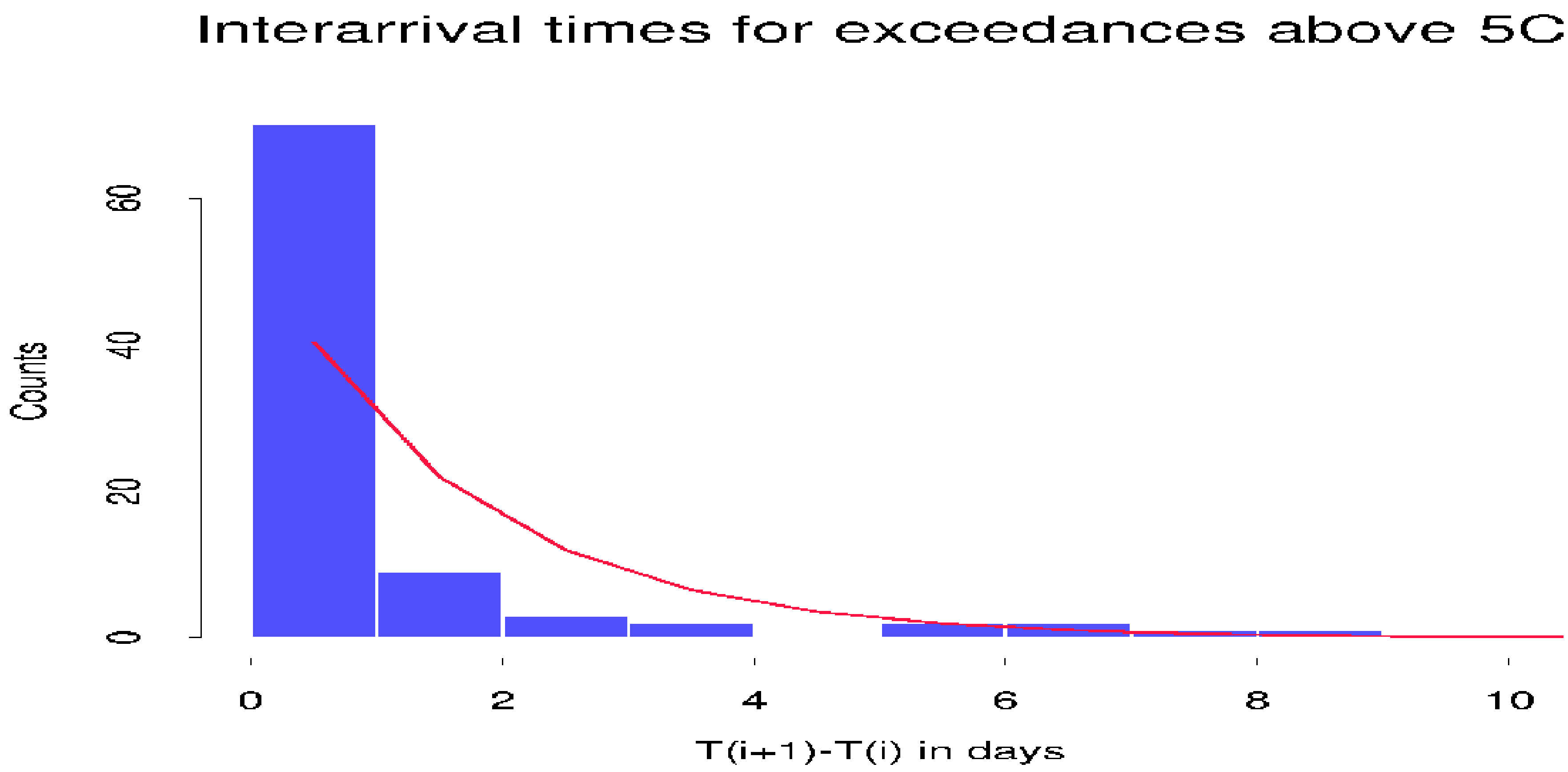
$$= \text{std. deviation(excess)/mean(excess)}$$



Note: values less than one imply Weibull type distribution with finite upper limit

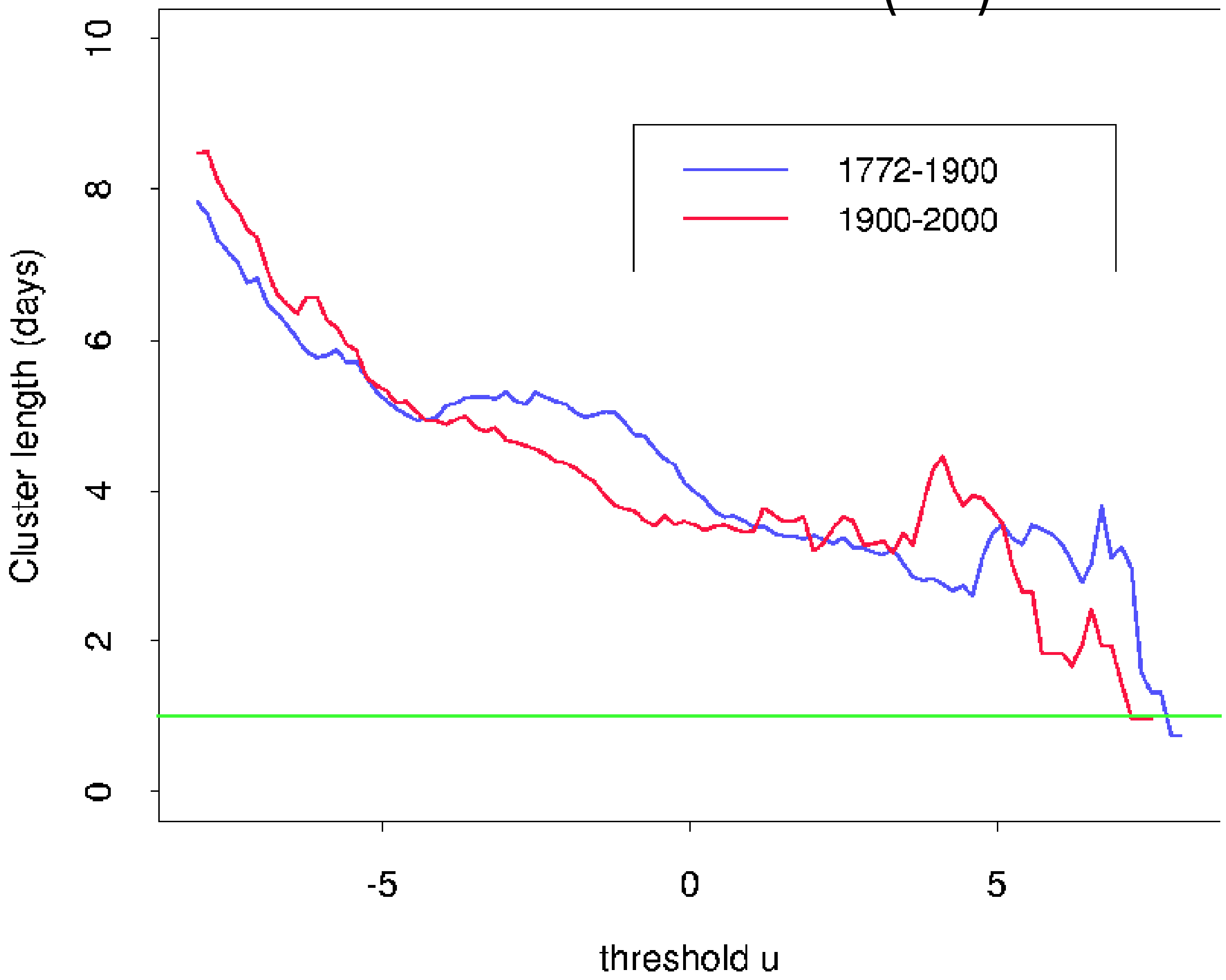
Clustering in time

For an unclustered point process the interarrival times T(i+1)-T(i) are exponentially distributed. However, extremes often come in clusters and it is important to take account of this dependency:



Changes in cluster length

$$L = (\text{Extremal Index})^{-1} = \frac{\overline{\Delta T^2}}{2\left(\overline{\Delta T}\right)^2}$$



Summary statistics

Period	R	Log R	M	V	L
1799-2000	0.091	-2.39	1.74	0.92	3.45
1799-1900	0.108	-2.23	1.82	0.93	3.98
1900-2000	0.070	-2.66	1.59	0.90	3.55
Change	-0.038	-0.43 Fewer extremes	-0.23 Less intense extremes	-0.03	-0.43 Shorter clusters

3. Understanding the changes …

Changes in location, scale, and shape

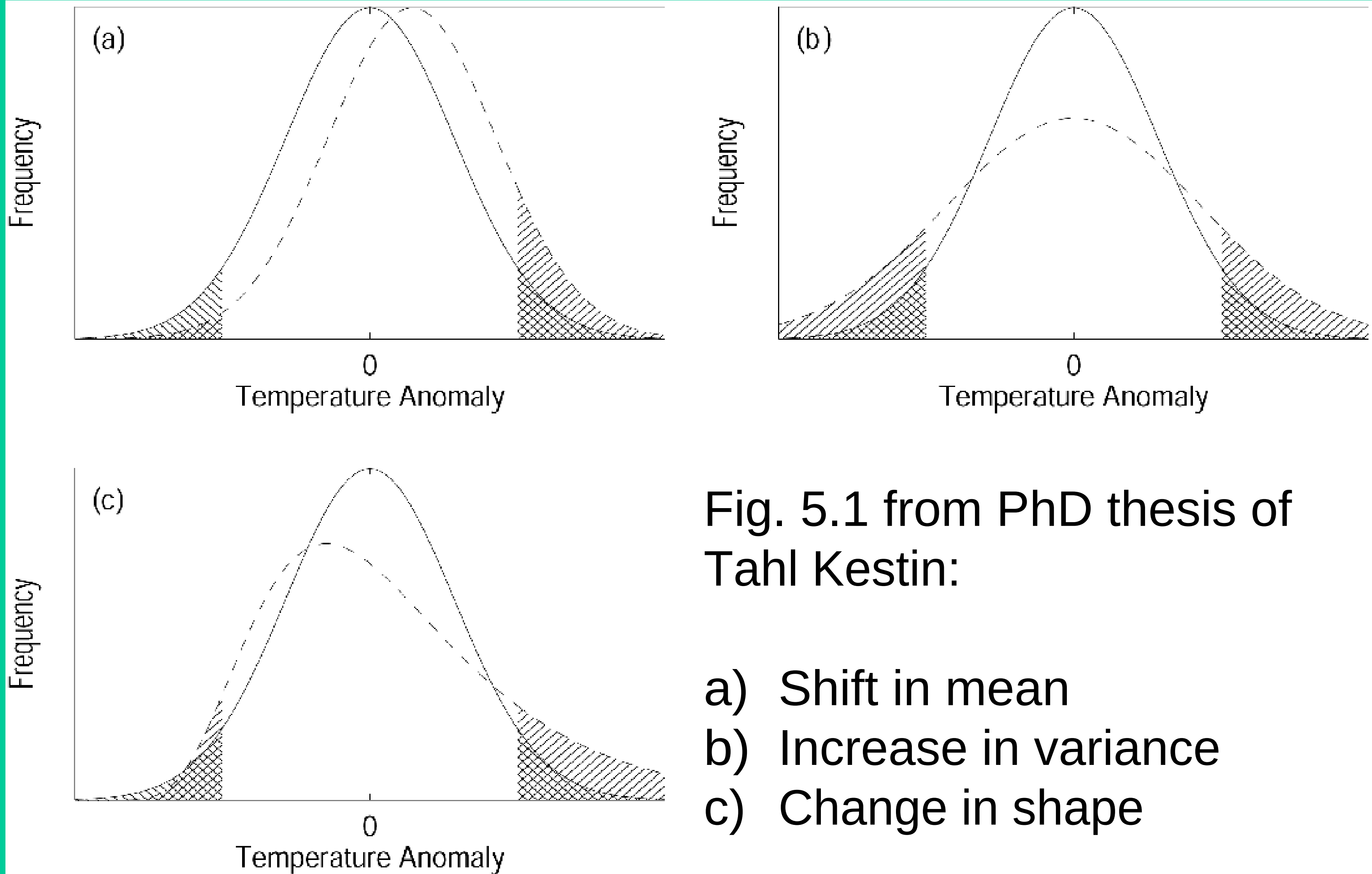


Fig. 5.1 from PhD thesis of Tahl Kestin:

a) Shift in mean
b) Increase in variance
c) Change in shape

Hypotheses about changing extremes

H0: **No change** (variation due to sampling only)

Sampling uncertainty can be tested using tail model

H1: Change due to **“mean effect”**

e.g. Mearns et al. (1984), Wigley (1985), …

H2: Change due to **“variance effect”**

e.g. Katz and Brown (1992), Katz and coworkers …

H3: Change due to **mean and variance effects**

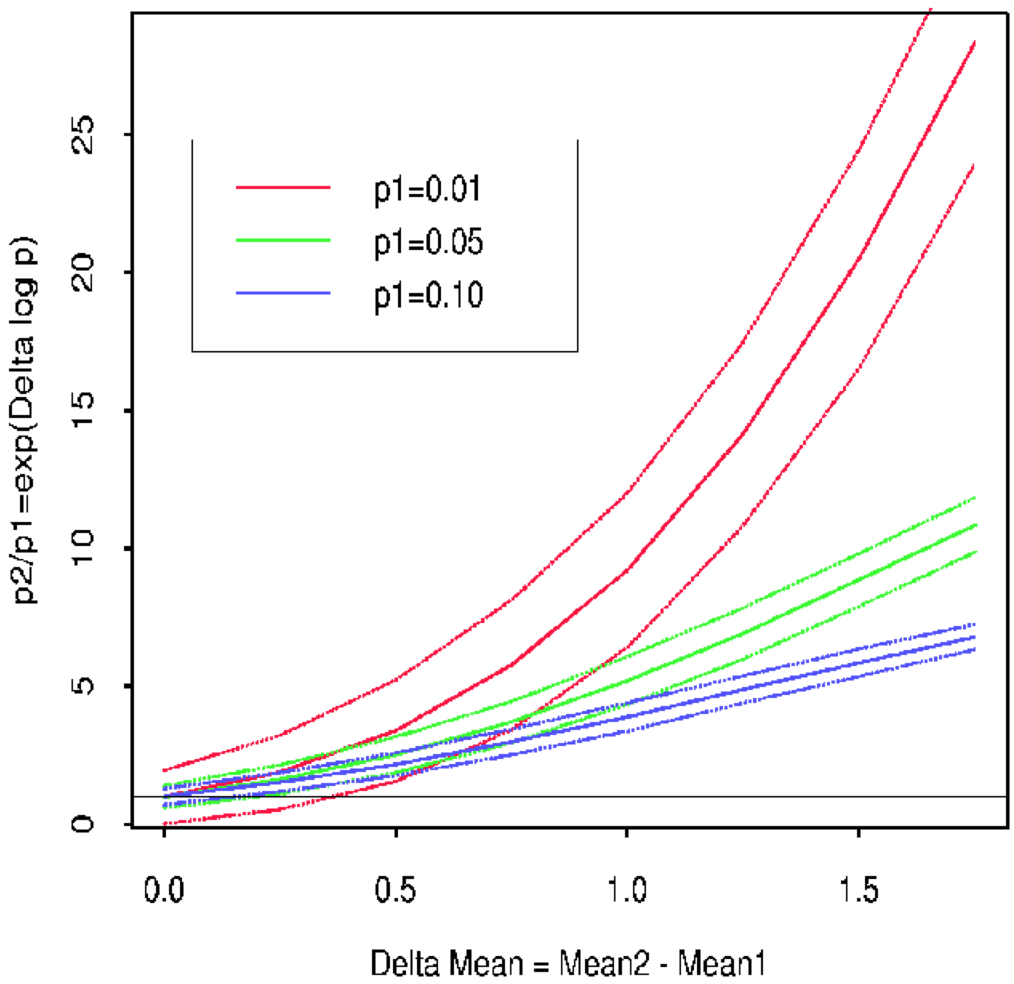
e.g. Brown and Katz (1995), …

H4: **“Structural change”** in shape etc

e.g. Kestin 2001, Antoniadou et al. 2001

Response to change in the mean

Change in exceedance rate for standard normal distribution
Dashed lines show 95% confidence intervals for estimates based on 400 samples

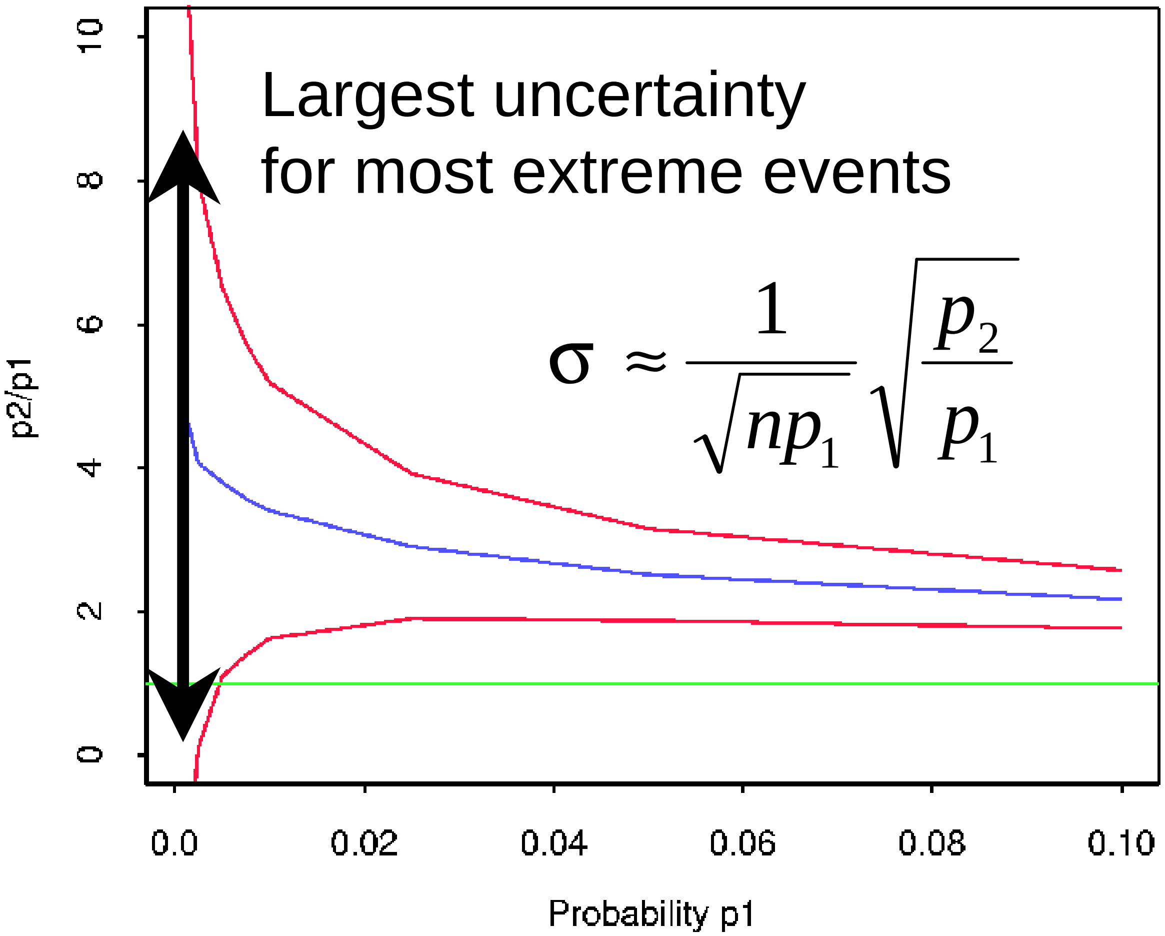


Note:

- nonlinear response in p2/p1
- p2/p1 is very sensitive especially for small p1
- sampling uncertainty also increases for small p1

Signal and noise for change in mean

Change in exceedance rate for shift of 0.5 in mean of the standard normal distribution. Red lines show 95% confidence intervalsfor estimates based on 400 samples

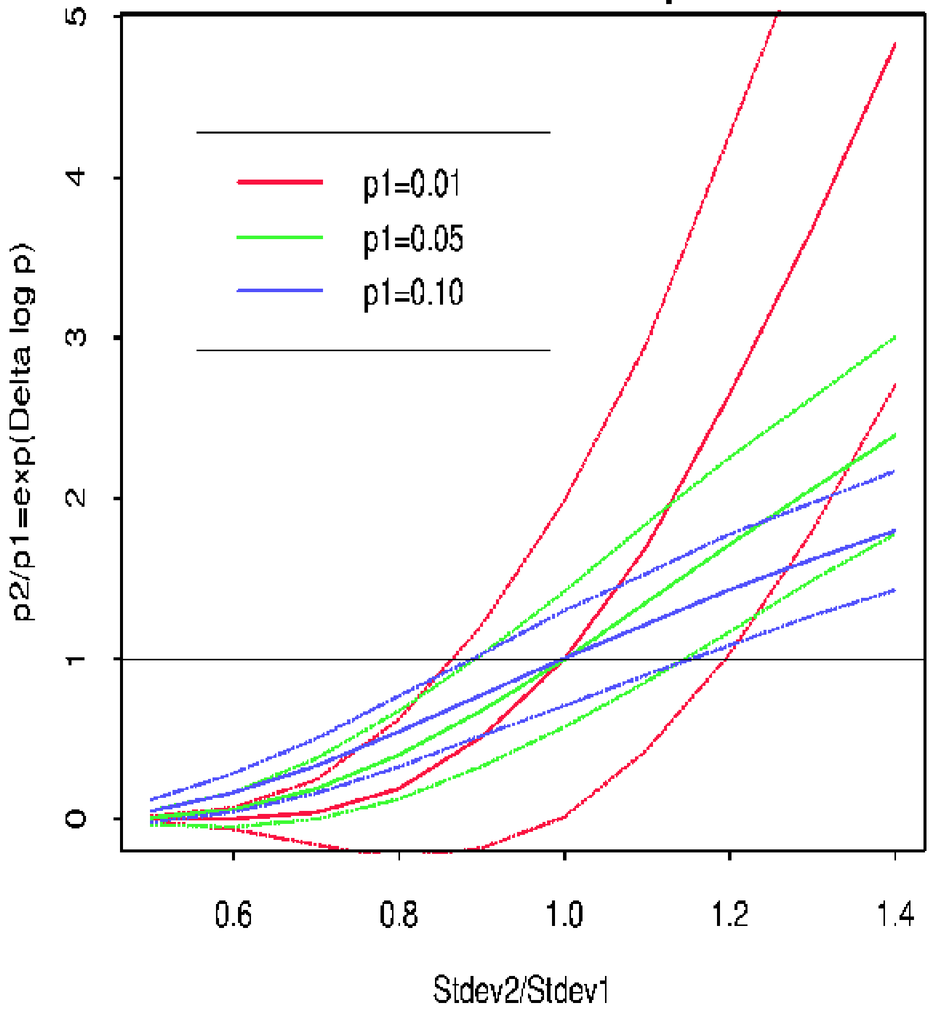


Note:

- largest change ("signal") for the most extreme events
- BUT also much more sampling uncertainty ("noise") in the tail
- signal/noise ratio decreases for more extreme events!

Response to change in variance

Change in exceedance rate for standard normal distribution
Dashed lines show 95% confidence intervalsfor estimates based on 400 samples



Note:

- very nonlinear response in p2/p1

- p2/p1 is very sensitive especially for small p1

- sampling uncertainty also increases substantially for small p1

Signal diagnosis

When $R = \Pr\{X > x\} = 1 - F\left(\frac{x-\mu}{\sigma}\right)$

$$\delta \log R = \left(\frac{f_x}{1-F_x}\right)\left\{\delta\mu + \left(\frac{x-\mu}{\sigma}\right)\delta\sigma\right\}$$

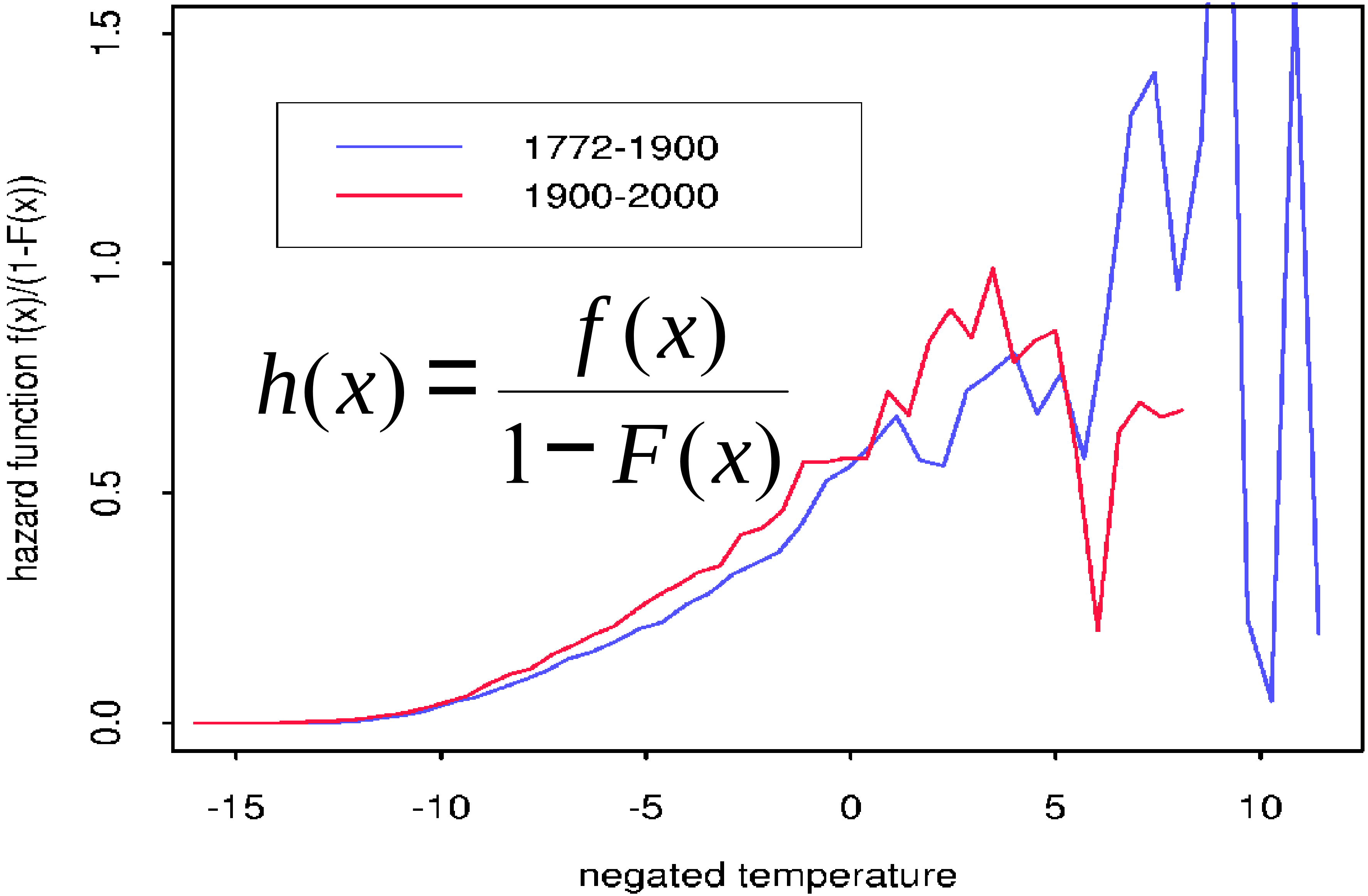
$$\Rightarrow p_2 / p_1 \approx e^{\{h(\delta\mu + z\delta\sigma)\}}$$

Example : for CET threshold x = 0 :

$d \log R = -0.43 = 0.5(-0.63 - 0.24)$

$\Rightarrow$ mean and variance change both contribute

The hazard function for CET



$$h(x) = \frac{f(x)}{1 - F(x)}$$

Structural change …

Antoniadou et al. 2001

L'OSCILLATION ATLANTIQUE NORD (NAO) ET INFLUENCE SUR LE CLIMAT EUROPÉEN 57

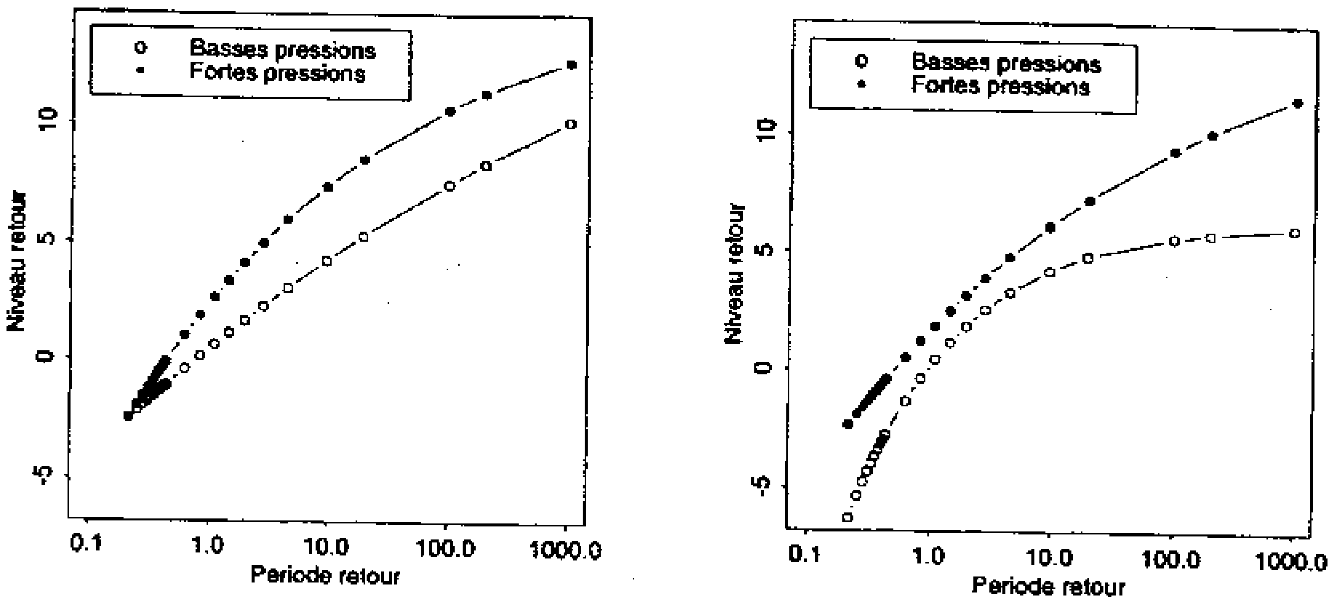


FIGURE 16

Graphe des estimations des niveaux retour pour les opposés des minima de températures hivernales en Angleterre, au 19ème siècle à gauche, et au 20ème siècle à droite, correspondant aux jours à «forte» ou à «faible» pression en Islande.

→ Tail shape parameter has become more sensitive to westerly flow NAO+ winters in the 20th century

4. Modelling the tail ...

Why do we need to model the tail?

- Quantify sampling uncertainty
- Make predictions beyond what is in the data sample e.g. larger values
- Use asymptotic theory to get the most information from the tail behaviour
- Summarise tail behaviour at ALL thresholds using just a few parameters

Note: Modelling whole of parent distribution can fail to capture proper tail behaviour e.g. rainfall extremes and gamma fits to parent.

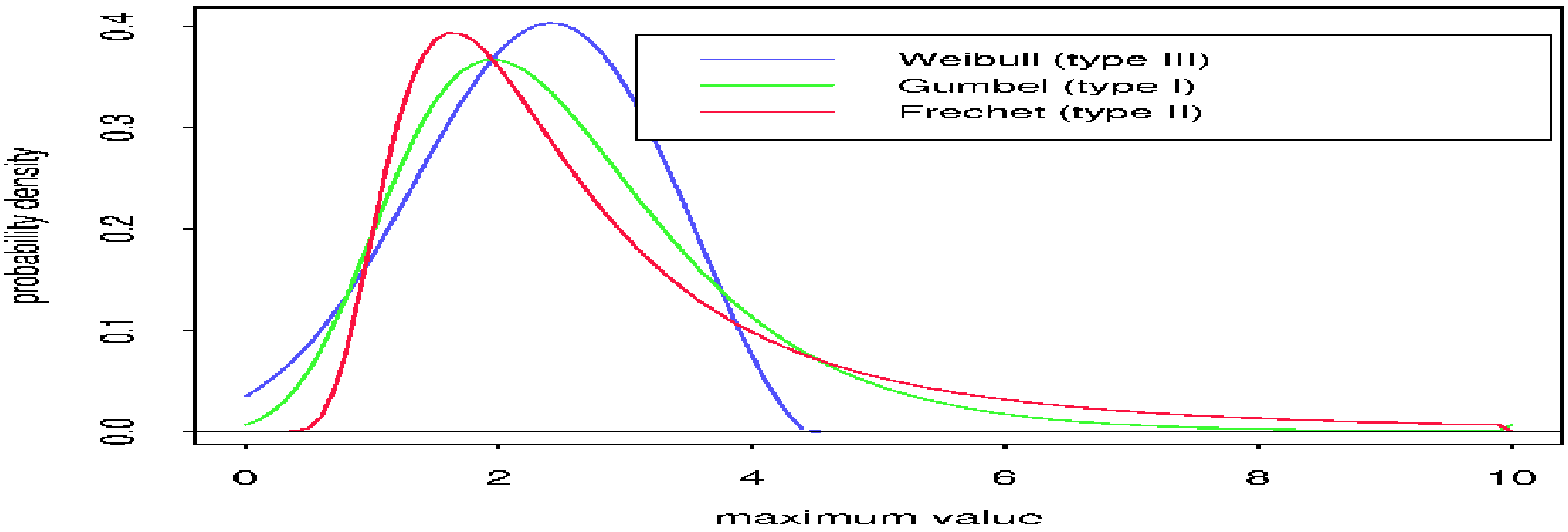
Tail parameter estimation methods …

1. **GEV fits to block maxima**
   - Uses only one value per block! (wastes data!)
   - Choice of block length?
2. **GEV fits to r-largest order statistics**
   - Choice of r?
   - Choice of block length?
   - Uses r values per block (better than maxima method)
3. **GPD fits to peaks over threshold**
   - Uses ALL data above threshold
   - Results should be insensitive to choice of threshold
   - Choice of threshold?
4. **Point process fits to peaks over threshold**
   - Uses ALL data above threshold
   - Stochastic process that can be generalised
   - Choice of threshold?

The Generalised Extreme Value distribution

Maxima/minima and the extreme tail of distributions can be modelled asymptotically by the 3-parameter Generalised Extreme Value (GEV) distribution:

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



Parametric description of attributes

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

$$M(u) = \frac{\sigma + \xi(u-\mu)}{1-\xi}$$

$$V(u) = \frac{1}{\sqrt{1-2\xi}}$$

Note: These expressions can be used to calculate the attributes for ANY threshold once the 3 tail parameters are known

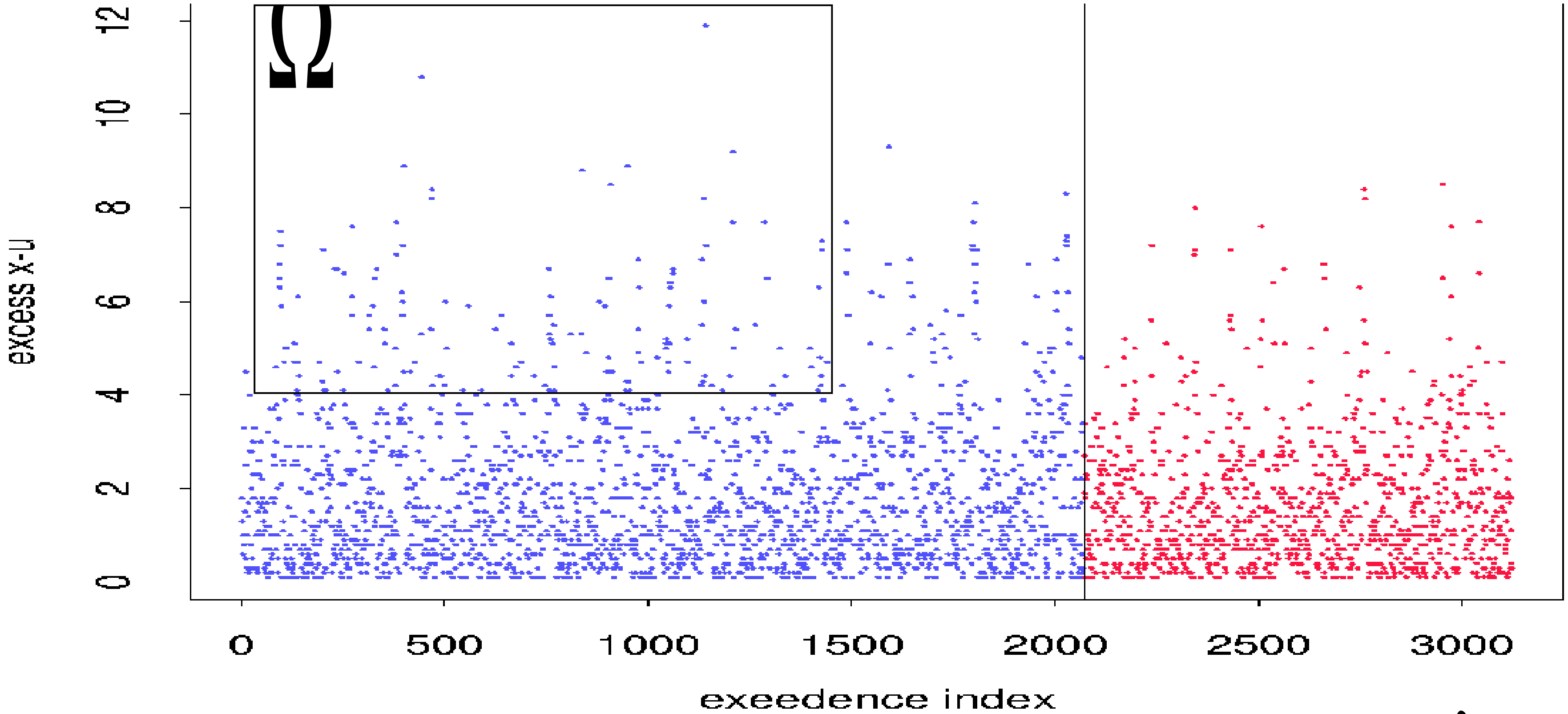
The Generalised Pareto Distribution

Excesses X-u above a high threshold u can be modelled using the 2-parameter Generalised Pareto Distribution:

$$\Pr\{\frac{X-u}{\sigma_u} > z\} = -(1+\xi z)^{-1/\xi}$$

where $\sigma_u = \sigma + \xi(u-\mu)$

Poisson process above high threshold



Probability of finding m points in region above x = $\dfrac{e^{-\lambda}\lambda^{m}}{m!}$

where rate $\lambda(\Omega) = (t_2 - t_1)\left[1 + \xi\left(\dfrac{x-u}{\sigma_u}\right)\right]^{-1/\xi}$

Point process estimates (with std. errors)

Period	μ	σ	ζ	$\mu - \sigma/\xi$
1799-2000	-5.70±0.22	2.75±0.15	-0.124±0.014	16.5
1799-1900	-5.47±0.26	2.81±0.18	-0.123±0.019	17.3
1900-2000	-6.20±0.43	2.82±0.27	-0.147±0.024	13.0
Change	-0.73(-2.03)	0.01(0.06)	-0.024 (-1.10)	-4.3

(*) are t values – change divided by standard error: $|t|>1.96$ is significant at 5% level

- statistically significant decrease in tail mean ($p<0.05$)
- hardly any change in tail variance
- large decrease in shape parameter (more convex)
- large reduction in upper limit of distribution

Conclusions ...

Main conclusions …

- Extremes are fascinating but require careful statistical analysis …
- ALL attributes of extremes should be investigated NOT just the rate of exceedence
- Tail parameters should be estimated using as much data as possible (e.g. not block maxima, use of pooling, etc.)
- Changes in attributes should be diagnosed in terms of changes in mean and variance of parent distribution (trends in variance=?)
- More complicated structural change might occur and can be diagnosed from regression on large-scale factors (covariates)

The End

For more information please refer to www.met.rdg.ac.uk/cag