RISK AND EXTREMES: ASSESSING THE PROBABILITIES OF VERY RARE EVENTS

Richard L. Smith
Department of Statistics and Operations Research
University of North Carolina
Chapel Hill, NC 27599-3260
rls@email.unc.edu

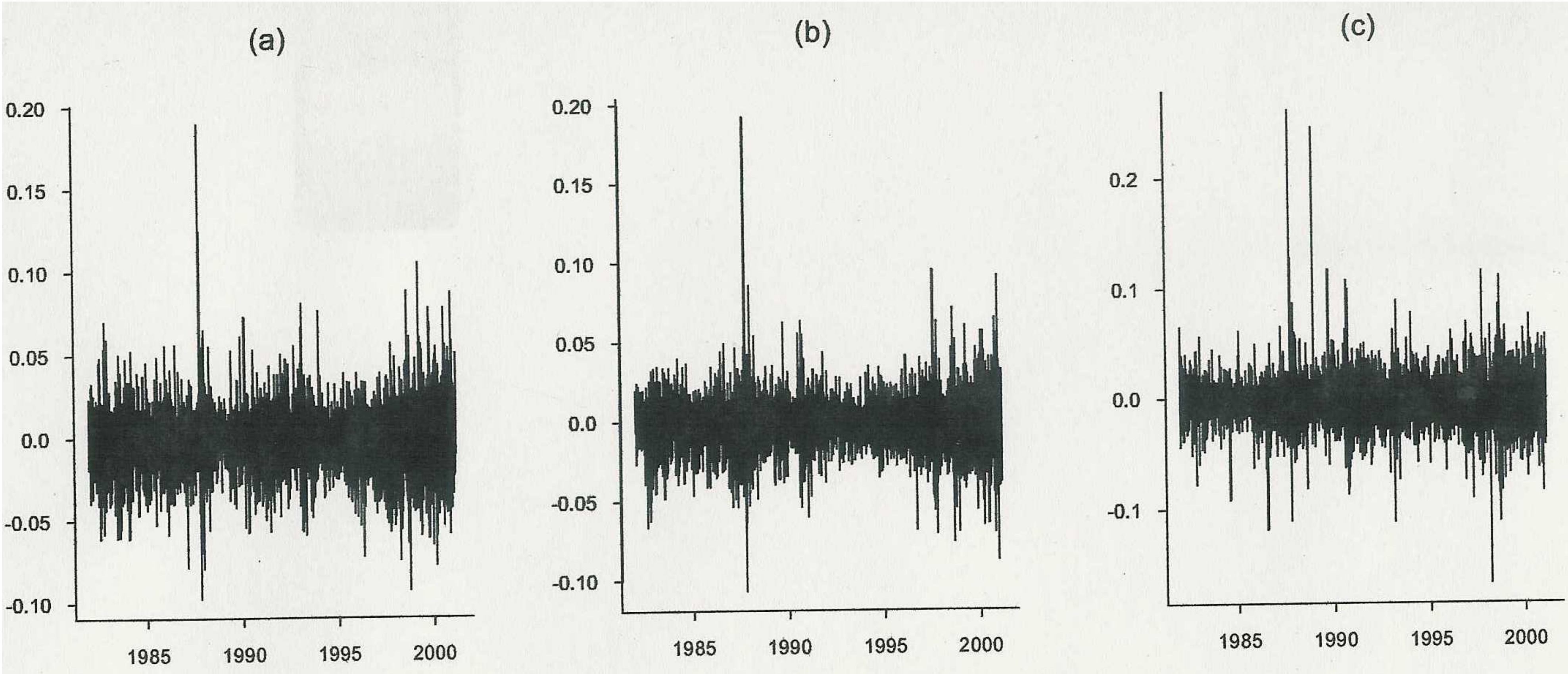
Fourteenth Army Conference on Applied Statistics
Virginia Military Institute
October 22, 2008

FAILED WIZARDS OF WALL STREET

Can you devise surefire ways to beat the markets? The rocket scientists thought they could. Boy, were they ever wrong

Smart people aren't supposed to get into this kind of a mess. With two Nobel prize winners among its partners, Long-Term Capital Management L.P. was considered too clever to get caught in a market downdraft. The Greenwich (Conn.) hedge fund nearly tripled the money of its wealthy investors between its inception in March, 1994, and the end of 1997. Its sophisticated arbitrage strategy was avowedly "market-neutral"--designed to make money whether prices were rising or falling. Indeed, until last spring its net asset value never fell more than 3% in a single month.

Then came the guns of August. Long-Term Capital's rocket science exploded on the launchpad. Its portfolio's value fell 44%, giving it a year-to-date decline of 52%. That's a loss of almost $2 billion. "August has been very painful for all of us," Chief Executive John W. Meriwether, a legendary bond trader, said in a letter to investors. (Long-Term's executives declined to speak on the record.)



Negative daily returns of Pfizer, GE and Citibank (thanks to Zhengjun Zhang)

These figures show negative daily returns from closing prices of 1982-2001 stock prices in three companies, Pfizer, GE and Citibank. Typical questions here are

1. How to determine *Value at Risk*, i.e. the amount which might be lost in a portfolio of assets over a specified time period with a specified small probability,

2. Dependence among the extremes of different series, and application to the portfolio management problem,

3. Modeling extremes in the presence of volatility.

INSURANCE EXTREMES

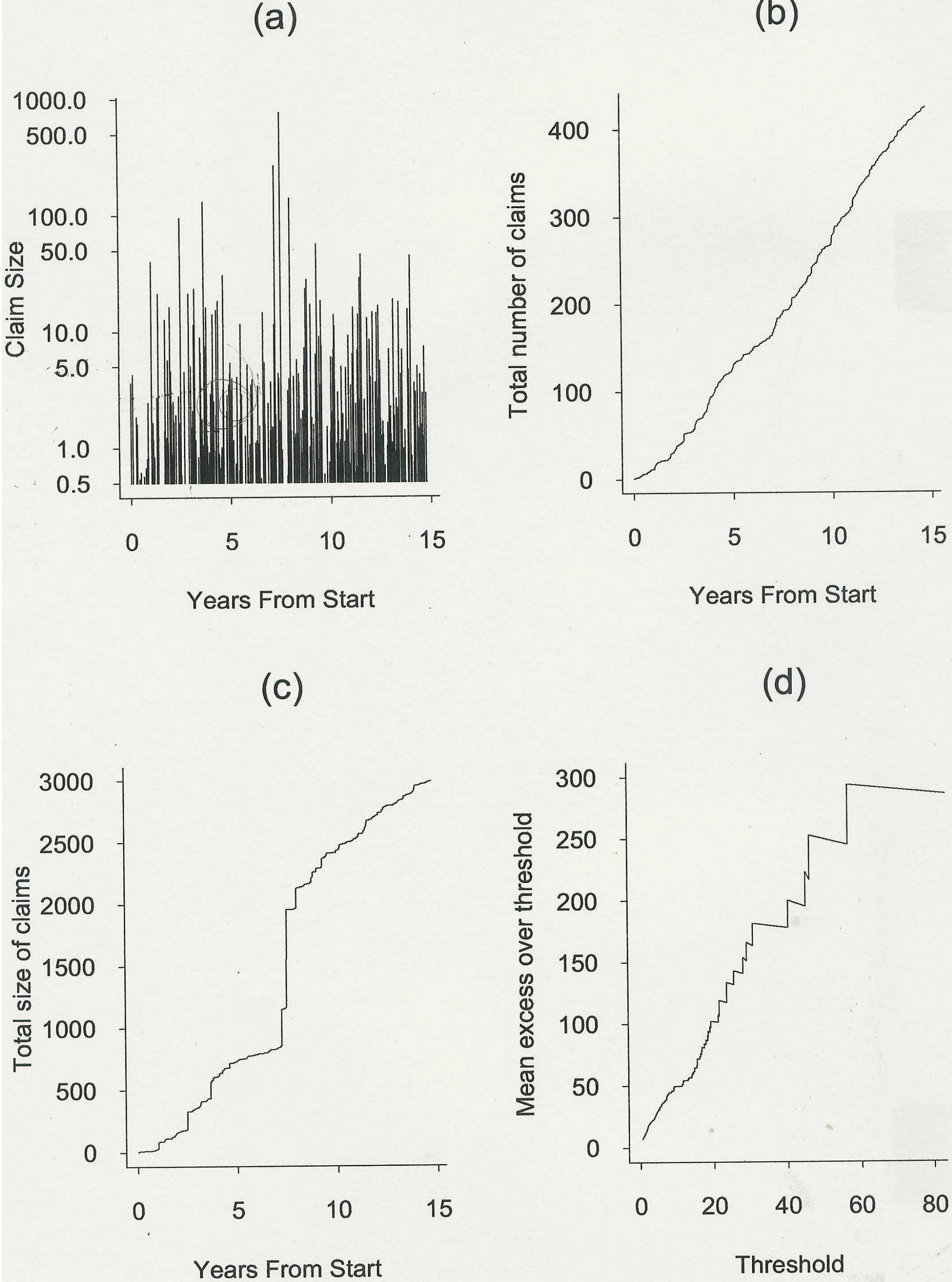
From Smith and Goodman (2000) —

The data consist of all insurance claims experienced by a large international oil company over a threshold 0.5 during a 15-year period — a total of 393 claims. Seven types:

Type	Description	Number	Mean
1	Fire	175	11.1
2	Liability	17	12.2
3	Offshore	40	9.4
4	Cargo	30	3.9
5	Hull	85	2.6
6	Onshore	44	2.7
7	Aviation	2	1.6

Total of all 393 claims: 2989.6

10 largest claims: 776.2, 268.0, 142.0, 131.0, 95.8, 56.8, 46.2, 45.2, 40.4, 30.7.



Some plots of the insurance data.

Some problems:

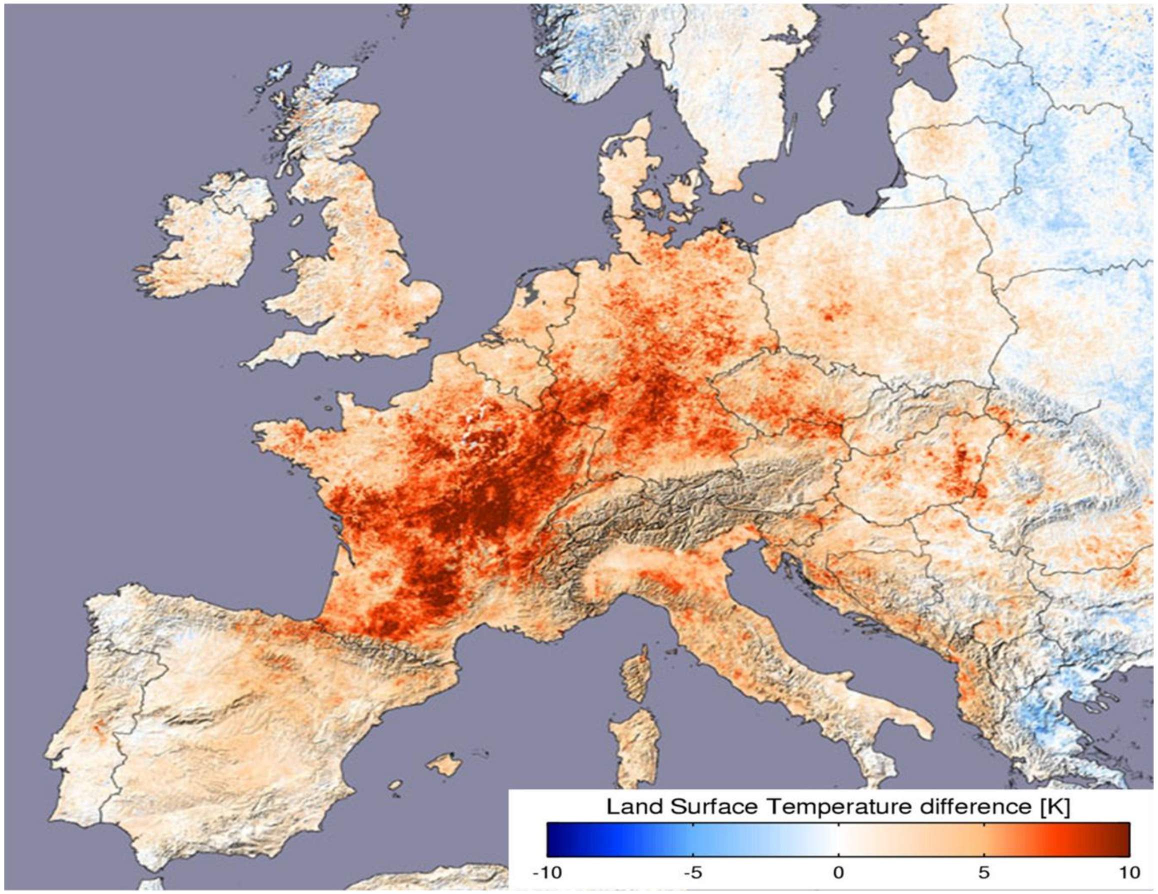
1. What is the distribution of very large claims?

2. Is there any evidence of a change of the distribution over time?

3. What is the influence of the different types of claim?

4. How should one characterize the risk to the company? More precisely, what probability distribution can one put on the amount of money that the company will have to pay out in settlement of large insurance claims over a future time period of, say, three years?

WEATHER EXTREMES



European temperatures in early August 2003, relative to 2001-2004 average

From NASA's MODIS - Moderate Resolution Imaging Spectrometer, courtesy of Reto Stöckli, ETHZ

(From a presentation by Myles Allen)

Human contribution to the European heatwave of 2003

Peter A. Stott[1], D. A. Stone[2,3] & M. R. Allen[2]

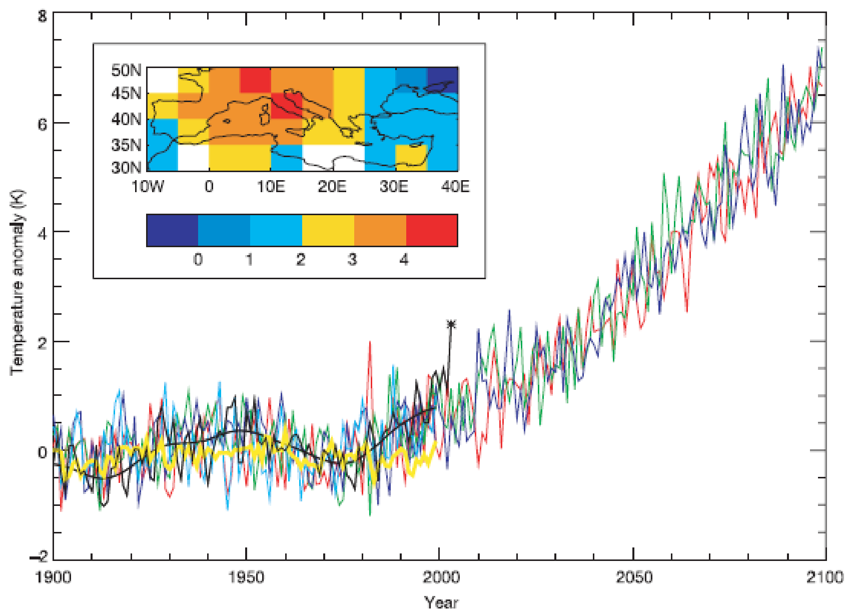


Figure 1 June–August temperature anomalies (relative to 1961–90 mean, in K) over the region shown in inset. Shown are observed temperatures (black line, with low-pass-filtered temperatures as heavy black line), modelled temperatures from four HadCM3 simulations including both anthropogenic and natural forcings to 2000 (red, green, blue and turquoise lines), and estimated HadCM3 response to purely natural natural forcings (yellow line). The observed 2003 temperature is shown as a star. Also shown (red, green and blue lines) are three simulations (initialized in 1989) including changes in greenhouse gas and sulphur emissions according to the SRES A2 scenario to 2100[22]. The inset shows observed summer 2003 temperature anomalies, in K.

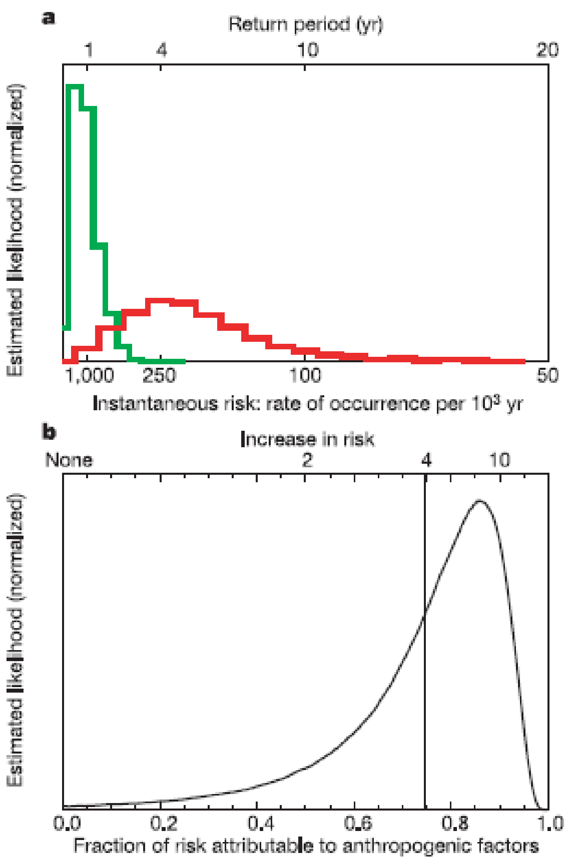


Figure 4 Change in risk of mean European summer temperatures exceeding the 1.6 K threshold. **a**, Histograms of instantaneous return periods under late-twentieth-century conditions in the absence of anthropogenic climate change (green line) and with anthropogenic climate change (red line). **b**, Fraction attributable risk (FAR). Also shown, as the vertical line, is the 'best estimate' FAR, the mean risk attributable to anthropogenic factors averaged over the distribution.

OUTLINE OF TALK

I. Extreme value theory

- Probability Models
- Estimation
- Diagnostics

II. Examples

III. Insurance Extremes

IV. Trends in Extreme Rainfall Events

I. EXTREME VALUE THEORY

EXTREME VALUE DISTRIBUTIONS

Suppose $X_1, X_2, ...,$ are independent random variables with the same probability distribution, and let $M_n = \max(X_1, ..., X_n)$. Under certain circumstances, it can be shown that there exist *normalizing constants* $a_n > 0, b_n$ such that

$$\Pr\left\{\frac{M_n - b_n}{a_n} \leq x\right\} = F(a_n x + b_n)^n \to H(x).$$

The *Three Types Theorem* (Fisher-Tippett, Gnedenko) asserts that if nondegenerate H exists, it must be one of three types:

$$\begin{aligned} H(x) &= \exp(-e^{-x}), \text{ all } x \quad \text{(Gumbel)} \\ H(x) &= \begin{cases} 0 & x < 0 \\ \exp(-x^{-\alpha}) & x > 0 \end{cases} \text{(Fréchet)} \\ H(x) &= \begin{cases} \exp(-|x|^{\alpha}) & x < 0 \\ 1 & x > 0 \end{cases} \text{(Weibull)} \end{aligned}$$

In Fréchet and Weibull, $\alpha > 0$.

The three types may be combined into a single *generalized extreme value* (GEV) distribution:

$$H(x) = \exp\left\{-\left(1+\xi\frac{x-\mu}{\psi}\right)_+^{-1/\xi}\right\},$$

$(y_+ = \max(y, 0))$

where μ is a location parameter, $\psi > 0$ is a scale parameter and ξ is a shape parameter. $\xi \to 0$ corresponds to the Gumbel distribution, $\xi > 0$ to the Fréchet distribution with $\alpha = 1/\xi$, $\xi < 0$ to the Weibull distribution with $\alpha = -1/\xi$.

$\xi > 0$: "long-tailed" case, $1 - F(x) \propto x^{-1/\xi}$,

$\xi = 0$: "exponential tail"

$\xi < 0$: "short-tailed" case, finite endpoint at $\mu - \xi/\psi$

EXCEEDANCES OVER THRESHOLDS

Consider the distribution of X conditionally on exceeding some high threshold u:

$$F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)}.$$

As $u \to \omega_F = \sup\{x : \ F(x) < 1\}$, often find a limit

$$F_u(y) \approx G(y; \sigma_u, \xi)$$

where G is *generalized Pareto distribution* (GPD)

$$G(y; \sigma, \xi) = 1 - \left(1 + \xi \frac{y}{\sigma}\right)_+^{-1/\xi}.$$

The Generalized Pareto Distribution

$$G(y;\sigma,\xi) = 1 - \left(1+\xi\frac{y}{\sigma}\right)_+^{-1/\xi}.$$

$\xi > 0$: long-tailed (equivalent to usual Pareto distribution), tail like $x^{-1/\xi}$,

$\xi = 0$: take limit as $\xi \to 0$ to get

$$G(y;\sigma,0) = 1 - \exp\left(-\frac{y}{\sigma}\right),$$

i.e. exponential distribution with mean σ,

$\xi < 0$: finite upper endpoint at $-\sigma/\xi$.

The *Poisson-GPD model* combines the GPD for the excesses over the threshold with a Poisson distribtion for the number of exceedances. Usually the mean of the Poisson distribution is taken to be λ per unit time.

OTHER MODELS FOR EXTREMES

1. *r-largest Order Statistics Model*: Instead of just considering the maximum value in each year, look at the r largest (from independent events), for some fixed small number r. Then we can also write down an asymptotic expression for the *joint* distribution of these r largest events. The case $r = 1$ reduces to classical extreme value theory.

2. There is also an alternative viewpoint known as the *point process approach*, which is a threshold approach like the GPD, but leads to a direct fitting of the GEV distribution for annual maxima. This can be very useful when we are interesting in the interrelations among the different approaches.

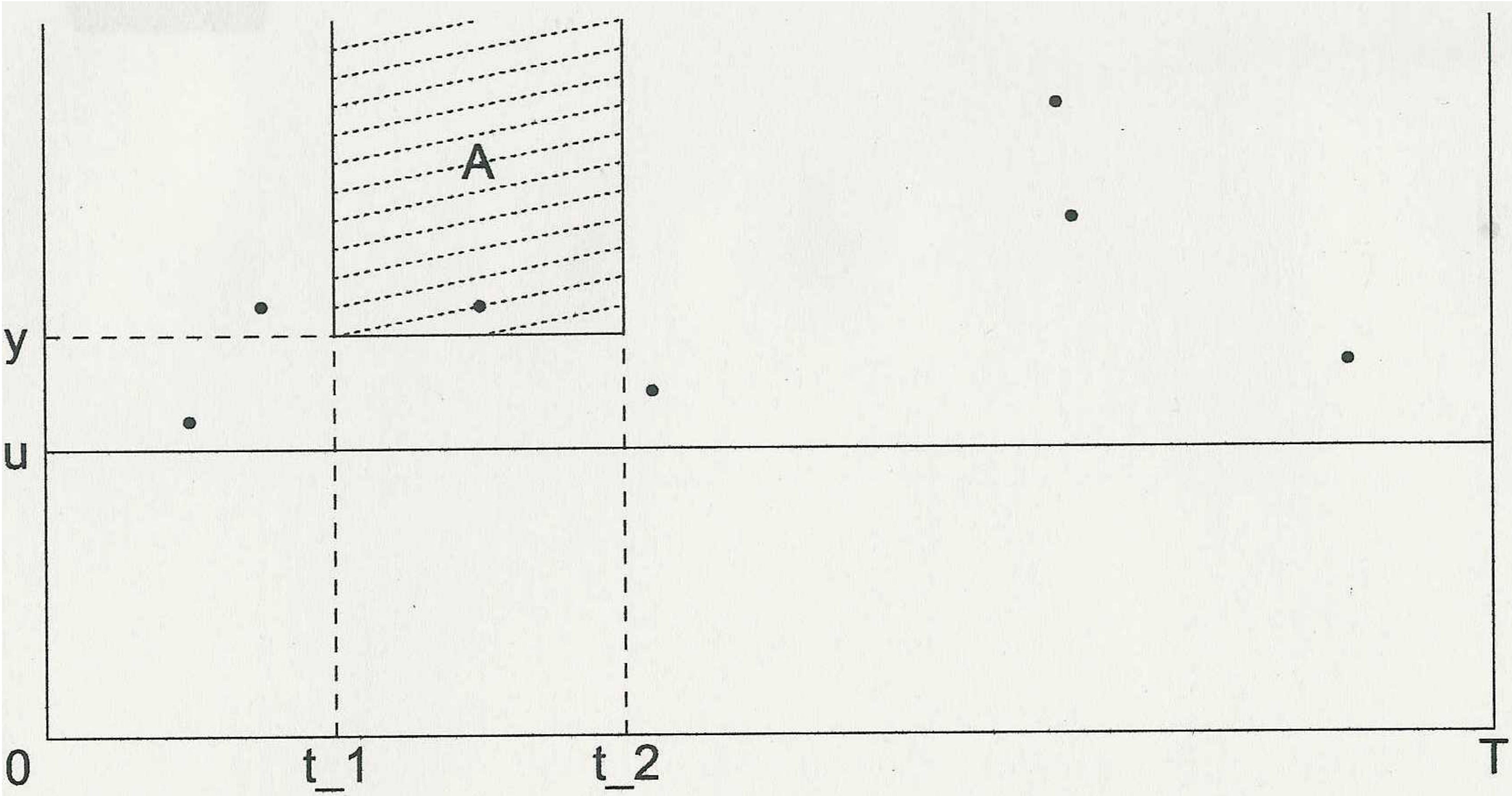


Illustration of point process model.

An extension of this approach allows for nonstationary processes in which the parameters μ, ψ and ξ are all allowed to be time-dependent, denoted μ_t, ψ_t and ξ_t.

This is the basis of the extreme value regression approaches introduced later

ESTIMATION

GEV log likelihood:

$$\ell = -N \log \psi - \left(\frac{1}{\xi} + 1\right) \sum_i \log \left(1 + \xi \frac{Y_i - \mu}{\psi}\right) - \sum_i \left(1 + \xi \frac{Y_i - \mu}{\psi}\right)^{-1/\xi}$$

provided $1 + \xi(Y_i - \mu)/\psi > 0$ for each i.

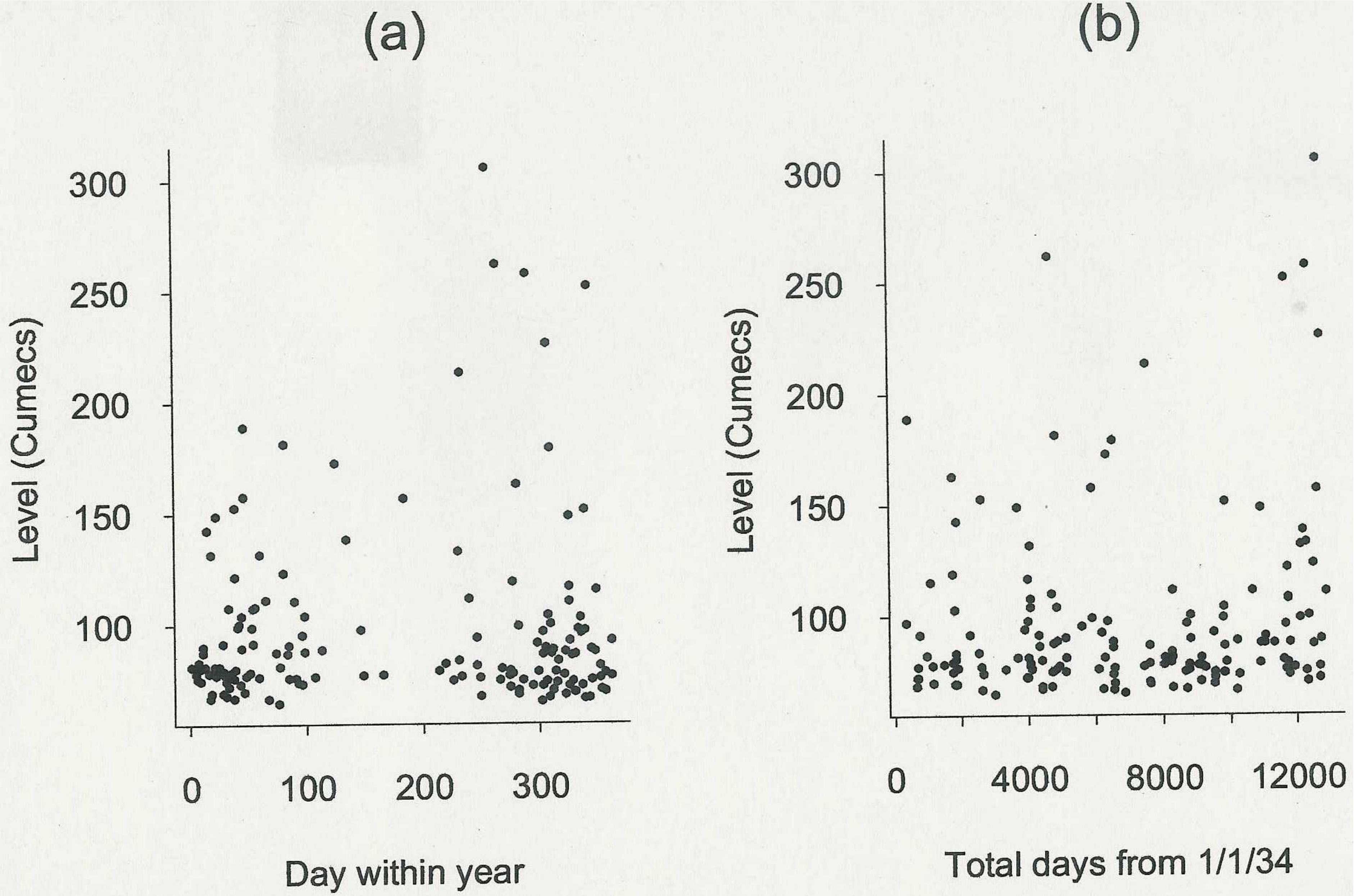
Poisson-GPD model:

$$\ell = N \log \lambda - \lambda T - N \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{N} \log \left(1 + \xi \frac{Y_i}{\sigma}\right)$$

provided $1 + \xi Y_i/\sigma > 0$ for all i.

The *method of maximum likelihood* states that we choose the parameters (μ, ψ, ξ) or (λ, σ, ξ) to maximize ℓ. These can be calculated numerically on the computer.

II. EXAMPLES



Plots of exceedances of River Nidd, (a) against day within year, (b) against total days from January 1, 1934. Adapted from Davison and Smith (1990).

DIAGNOSTICS

Gumbel plots

QQ plots of residuals

Mean excess plot

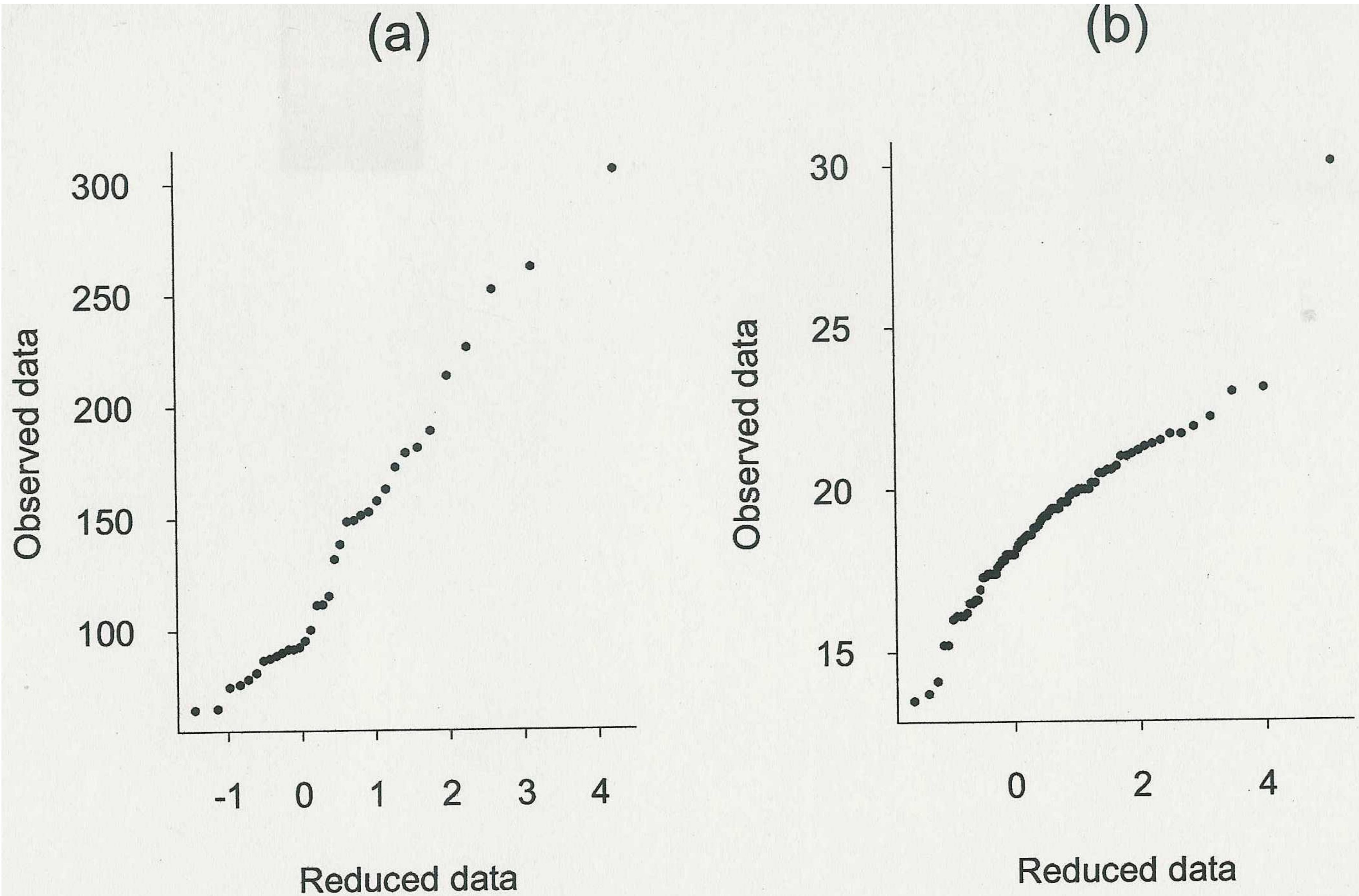
Gumbel plots

Used as a diagnostic for Gumbel distribution with annual maxima data. Order data as $Y_{1:N} \le ... \le Y_{N:N}$, then plot $Y_{i:N}$ against *reduced value* $x_{i:N}$,

$$x_{i:N} = -\log(-\log p_{i:N}),$$

$p_{i:N}$ being the i'th *plotting position*, usually taken to be $(i-\frac{1}{2})/N$.

A straight line is ideal. Curvature may indicate Fréchet or Weibull form. Also look for outliers.



Gumbel plots. (a) Annual maxima for River Nidd flow series. (b) Annual maximum temperatures in Ivigtut, Iceland.

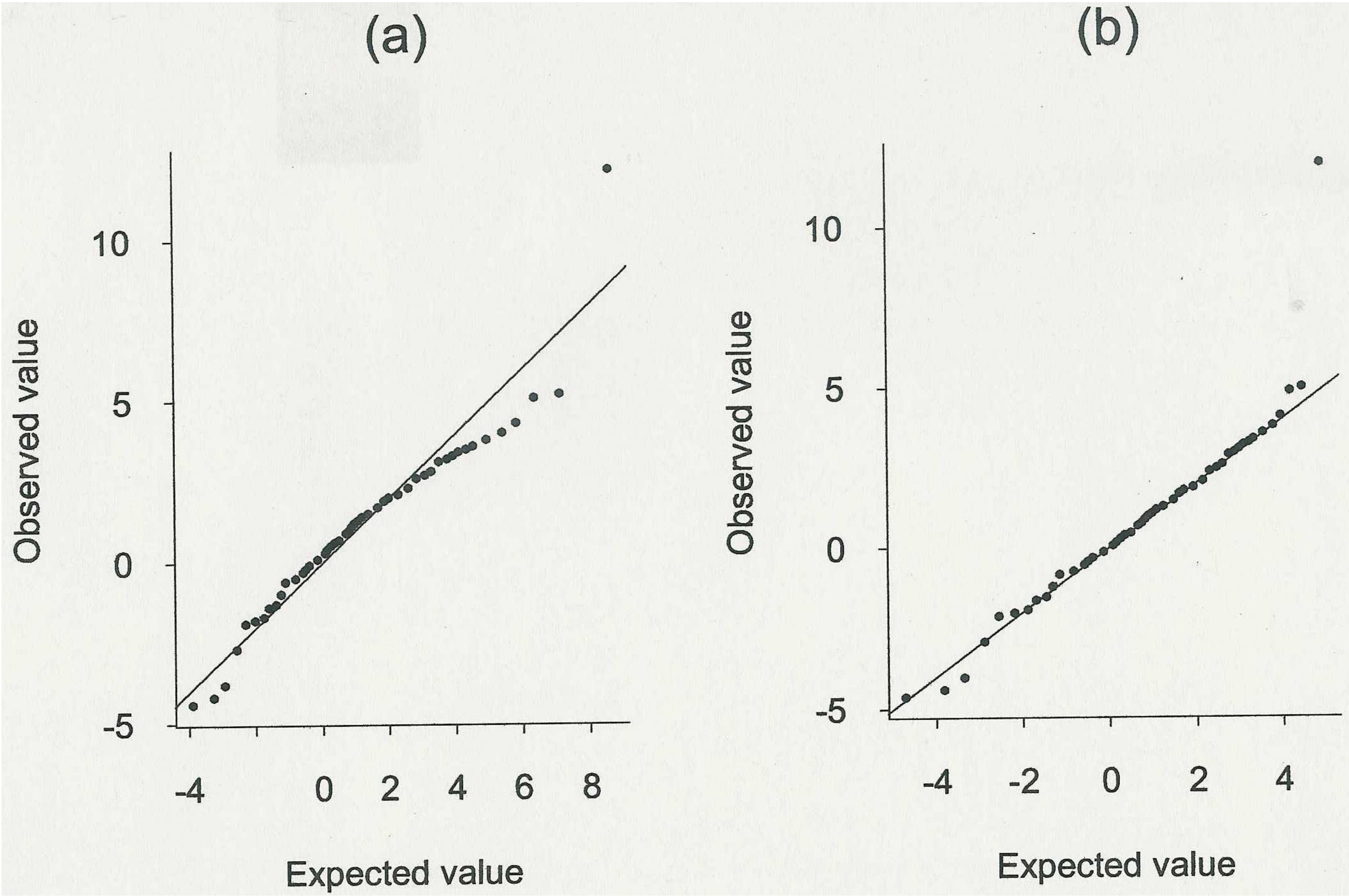
QQ plots of residuals

A second type of probability plot is drawn *after* fitting the model. Suppose $Y_1, ..., Y_N$ are IID observations whose common distribution function is $G(y;\theta)$ depending on parameter vector θ. Suppose θ has been estimated by $\widehat{\theta}$, and let $G^{-1}(p;\theta)$ denote the inverse distribution function of G, written as a function of θ. A QQ (quantile-quantile) plot consists of first ordering the observations $Y_{1:N} \le ... \le Y_{N:N}$, and then plotting $Y_{i:N}$ against the reduced value

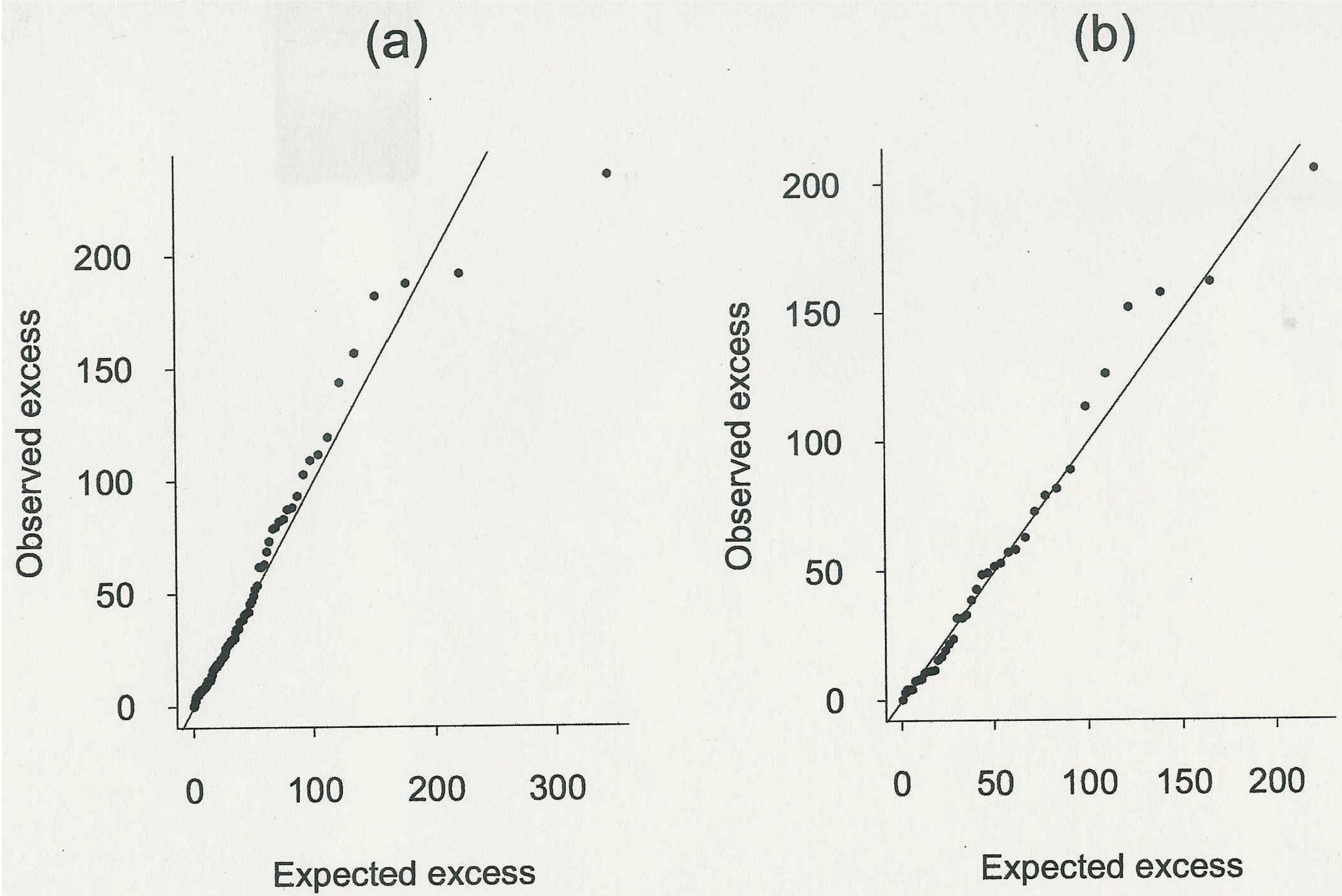
$$x_{i:N} = G^{-1}(p_{i:N};\widehat{\theta}),$$

where $p_{i:N}$ may be taken as $(i-\frac{1}{2})/N$. If the model is a good fit, the plot should be roughly a straight line of unit slope through the origin.

Examples...



GEV model to Ivigtut data, (a) without adjustment, (b) excluding largest value from model fit but including it in the plot.



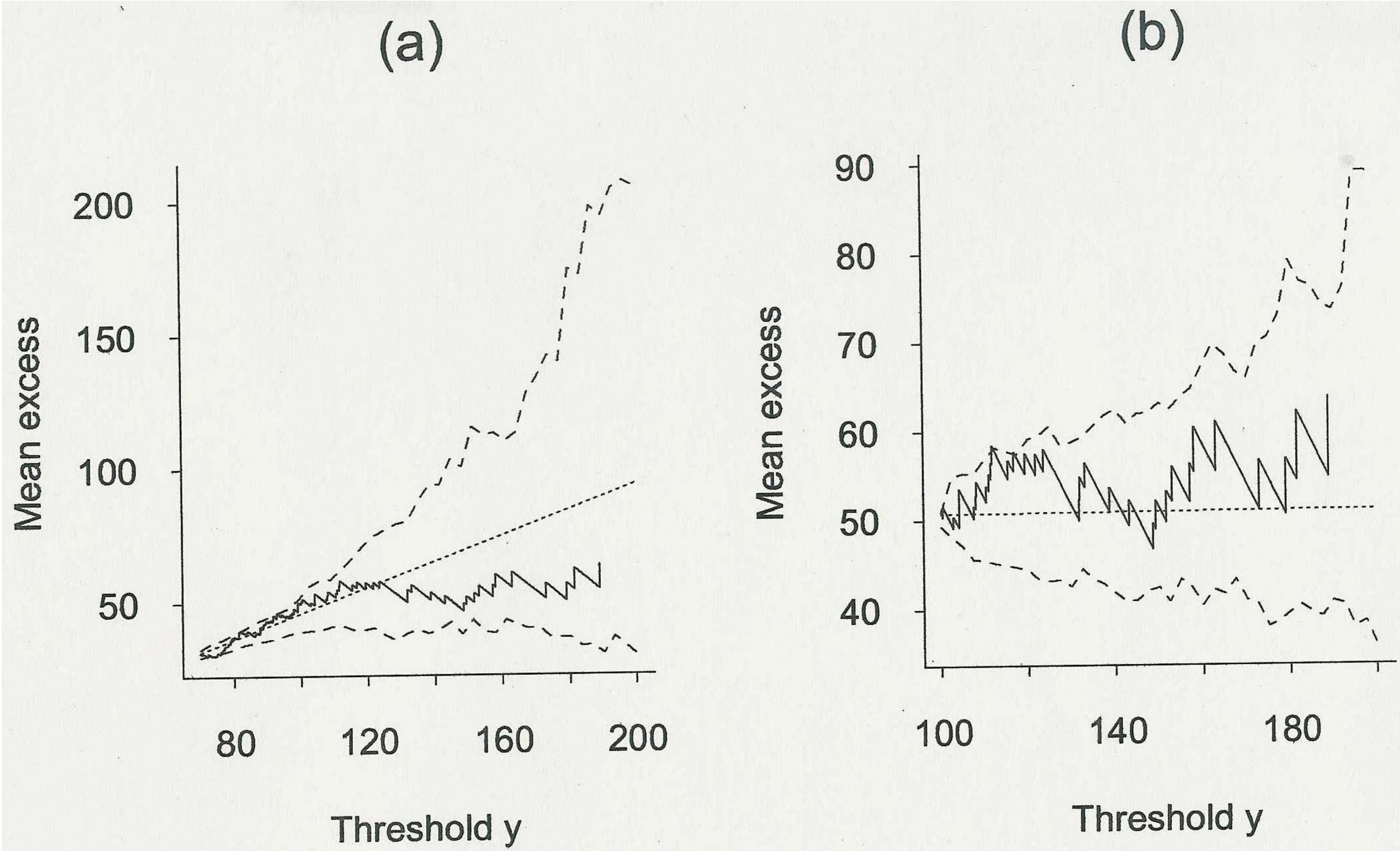
QQ plots for GPD, Nidd data. (a) $u = 70$. (b) $u = 100$.

Mean excess plot

Idea: for a sequence of values of w, plot the mean excess over w against w itself. If the GPD is a good fit, the plot should be approximately a straight line.

In practice, the actual plot is very jagged and therefore its "straightness" is difficult to assess. However, a Monte Carlo technique, *assuming* the GPD is valid throughout the range of the plot, can be used to assess this.

Examples...



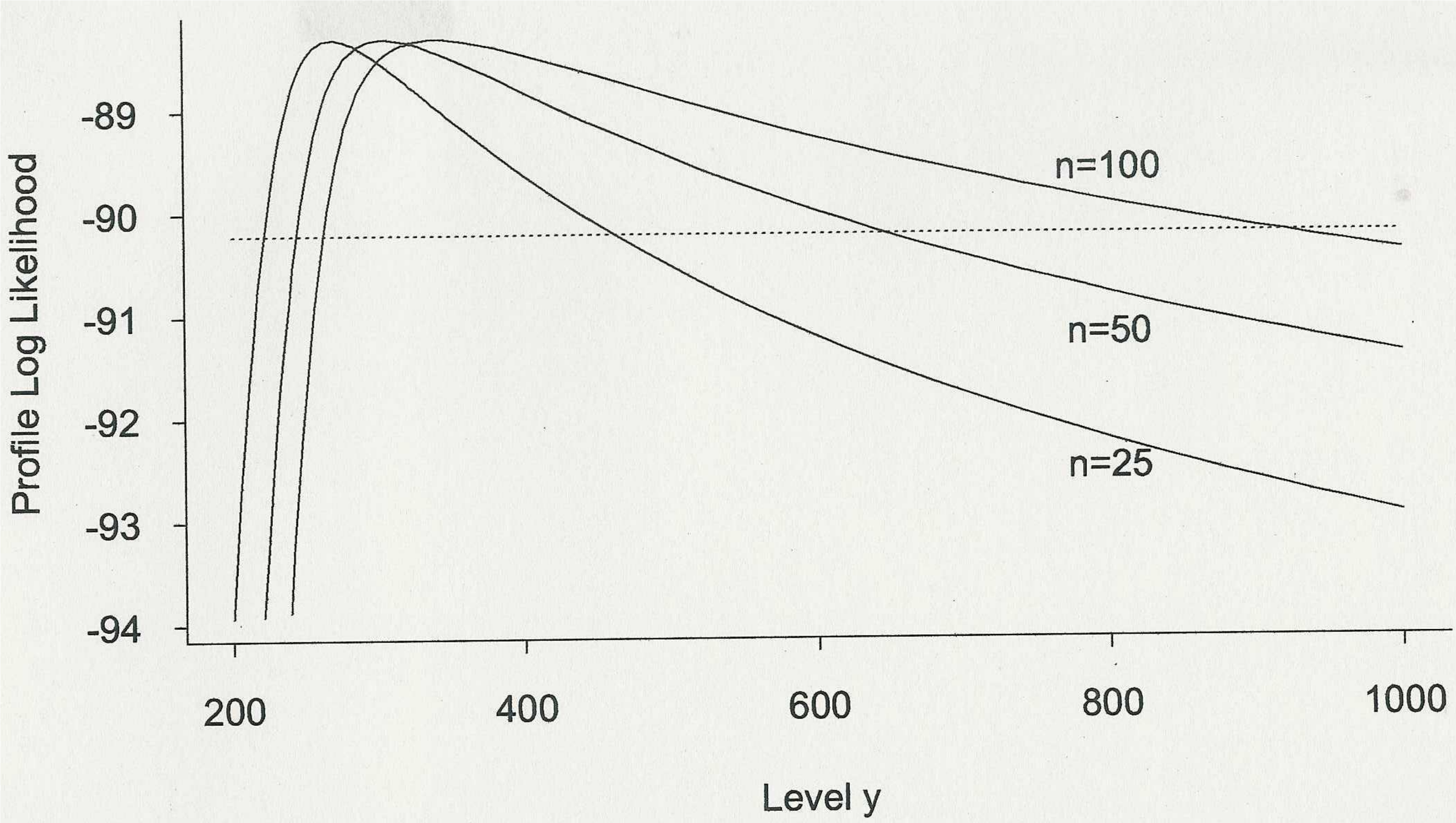
Mean excess over threshold plots for Nidd data, with Monte Carlo confidence bands, relative to threshold 70 (a) and 100 (b).

PROFILES LIKELIHOODS FOR QUANTILES

Suppose we are interested in the N-year return level y_N, i.e. the $(1-1/N)$-quantile of the annual maximum distribution. We can express y_N as a function of the extreme value parameters μ, ψ and ξ, and thereby obtain an estimate for any N.

However, that raises the question of what is the uncertainty of this estimate. A very general approach to this is via something called the *profile likelihood*, which calculates the likelihood of y_N after maximizing with respect to the other parameters.

Example from the Nidd data:



Profile log-likelihoods for extreme quantiles based on Nidd data

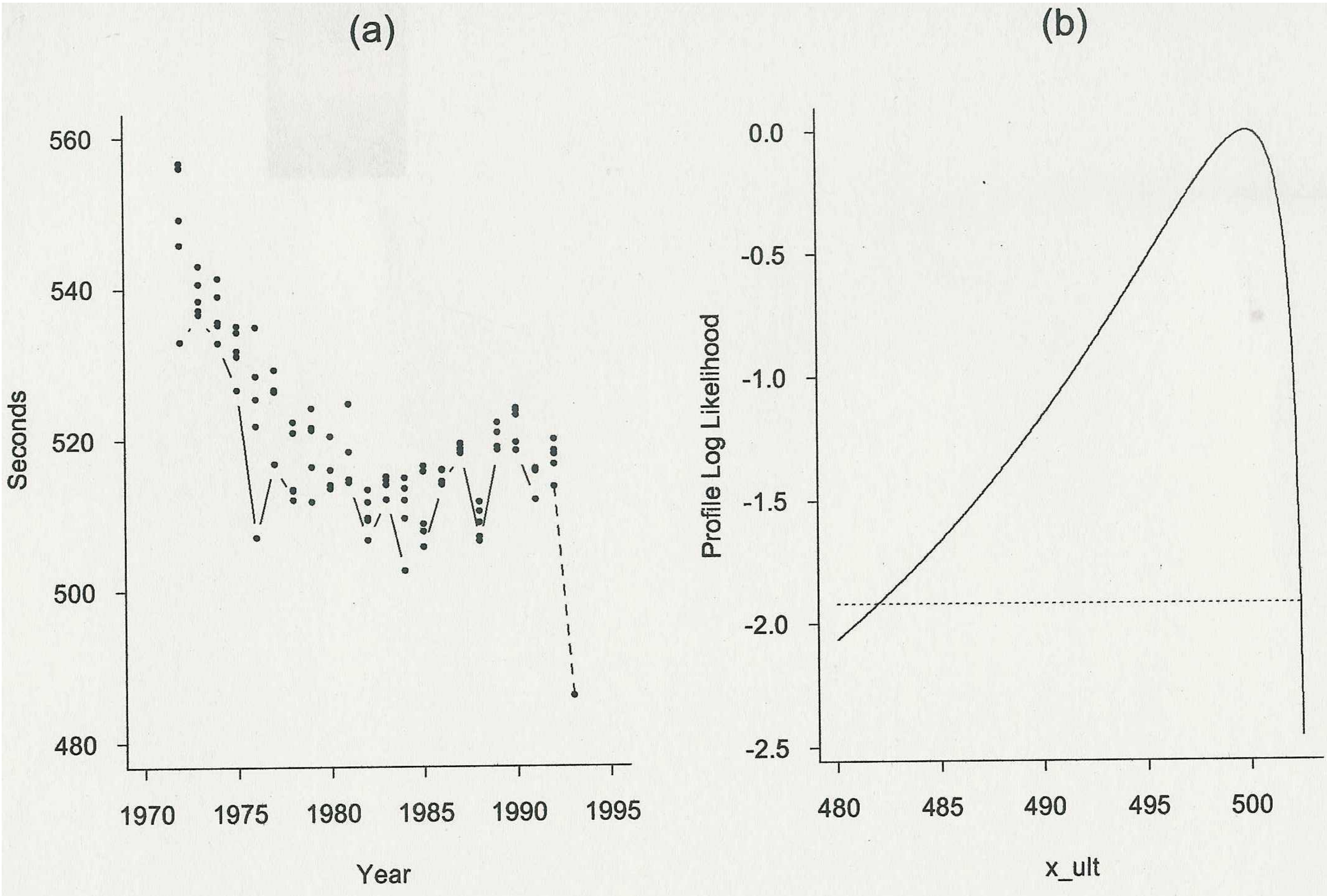
BAYESIAN APPROACHES

An alternative approach to extreme value inference is Bayesian, using vague priors for the GEV parameters and MCMC samples for the computations. Bayesian methods are particularly useful for *predictive inference*, e.g. if Z is some as yet unobserved random variable whose distribution depends on μ, ψ and ξ, estimate $\text{Pr}\{Z > z\}$ by

$$\int \text{Pr}\{Z > z; \mu, \psi, \xi\} \pi(\mu, \psi, \xi | Y) d\mu d\psi d\xi$$

where $\pi(...|Y)$ denotes the posterior density given past data Y


ITALIA
Atlanta 1996
3438
Atlanta 1996
3154



Plots of women's 3000 meter records, and profile log-likelihood for ultimate best value based on pre-1993 data.

Example. The left figure shows the five best running times by different athletes in the women's 3000 metre track event for each year from 1972 to 1992. Also shown on the plot is Wang Junxia's world record from 1993. Many questions were raised about possible illegal drug use.

We approach this by asking how implausible Wang's performance was, given all data up to 1992.

Robinson and Tawn (1995) used the r largest order statistics method (with $r = 5$, translated to smallest order statistics) to estimate an extreme value distribution, and hence computed a profile likelihood for x_{ult}, the lower endpoint of the distribution, based on data up to 1992 (right plot of previous figure)

Alternative Bayesian calculation:

(Smith 1997)

Compute the (Bayesian) predictive probability that the 1993 performance is equal or better to Wang's, given the data up to 1992, and conditional on the event that there is a new world record.

The answer is approximately 0.0006.

III. INSURANCE EXTREMES

From Smith and Goodman (2000) —

The data consist of all insurance claims experienced by a large international oil company over a threshold 0.5 during a 15-year period — a total of 393 claims. Seven types:

Type	Description	Number	Mean
1	Fire	175	11.1
2	Liability	17	12.2
3	Offshore	40	9.4
4	Cargo	30	3.9
5	Hull	85	2.6
6	Onshore	44	2.7
7	Aviation	2	1.6

Total of all 393 claims: 2989.6

10 largest claims: 776.2, 268.0, 142.0, 131.0, 95.8, 56.8, 46.2, 45.2, 40.4, 30.7.

GPD fits to various thresholds:

u	N_u	Mean Excess	σ	ξ
0.5	393	7.11	1.02	1.01
2.5	132	17.89	3.47	0.91
5	73	28.9	6.26	0.89
10	42	44.05	10.51	0.84
15	31	53.60	5.68	1.44
20	17	91.21	19.92	1.10
25	13	113.7	74.46	0.93
50	6	37.97	150.8	0.29

Point process approach:

u	N_u	μ	$\log\psi$	ξ
0.5	393	26.5 (4.4)	3.30 (0.24)	1.00 (0.09)
2.5	132	26.3 (5.2)	3.22 (0.31)	0.91 (0.16)
5	73	26.8 (5.5)	3.25 (0.31)	0.89 (0.21)
10	42	27.2 (5.7)	3.22 (0.32)	0.84 (0.25)
15	31	22.3 (3.9)	2.79 (0.46)	1.44 (0.45)
20	17	22.7 (5.7)	3.13 (0.56)	1.10 (0.53)
25	13	20.5 (8.6)	3.39 (0.66)	0.93 (0.56)

Standard errors are in parentheses

Predictive Distributions of Future Losses

What is the probability distribution of future losses over a specific time period, say 1 year?

Let Y be future total loss. Distribution function $G(y; \mu, \psi, \xi)$ — in practice this must itself be simulated.

Traditional frequentist approach:

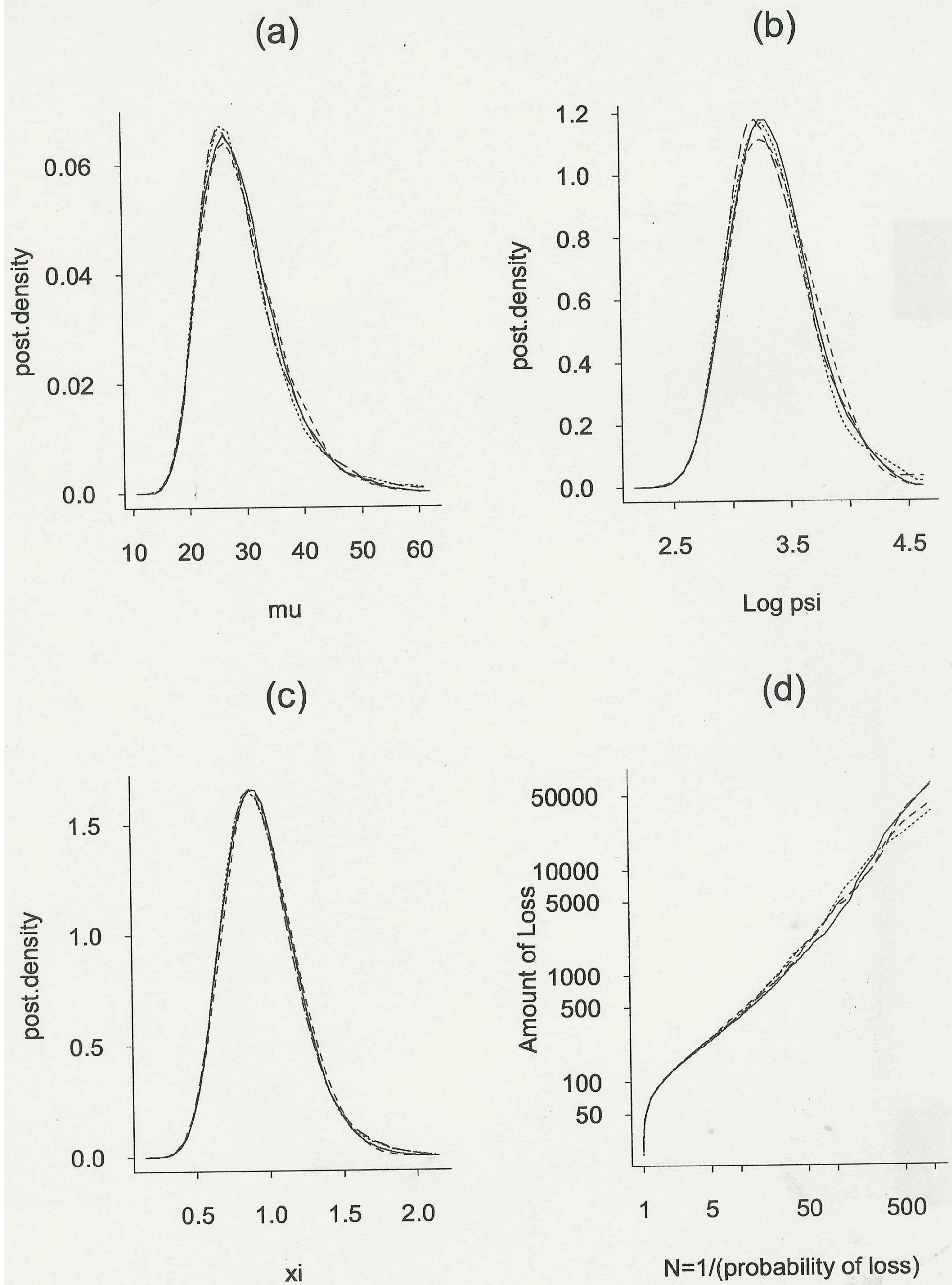
$$\widehat{G}(y) = G(y; \widehat{\mu}, \widehat{\psi}, \widehat{\xi})$$

where $\widehat{\mu}$, $\widehat{\psi}$, $\widehat{\xi}$ are MLEs.

Bayesian:

$$\tilde{G}(y) = \int G(y; \mu, \psi, \xi) d\pi(\mu, \psi, \xi \mid \mathbf{X})$$

where $\pi(\cdot \mid \mathbf{X})$ denotes posterior density given data $\mathbf{X}$.



Estimated posterior densities for the three parameters, and for the predictive distribution function. Four independent Monte Carlo runs are shown for each plot.

Hierarchical models for claim type and year effects

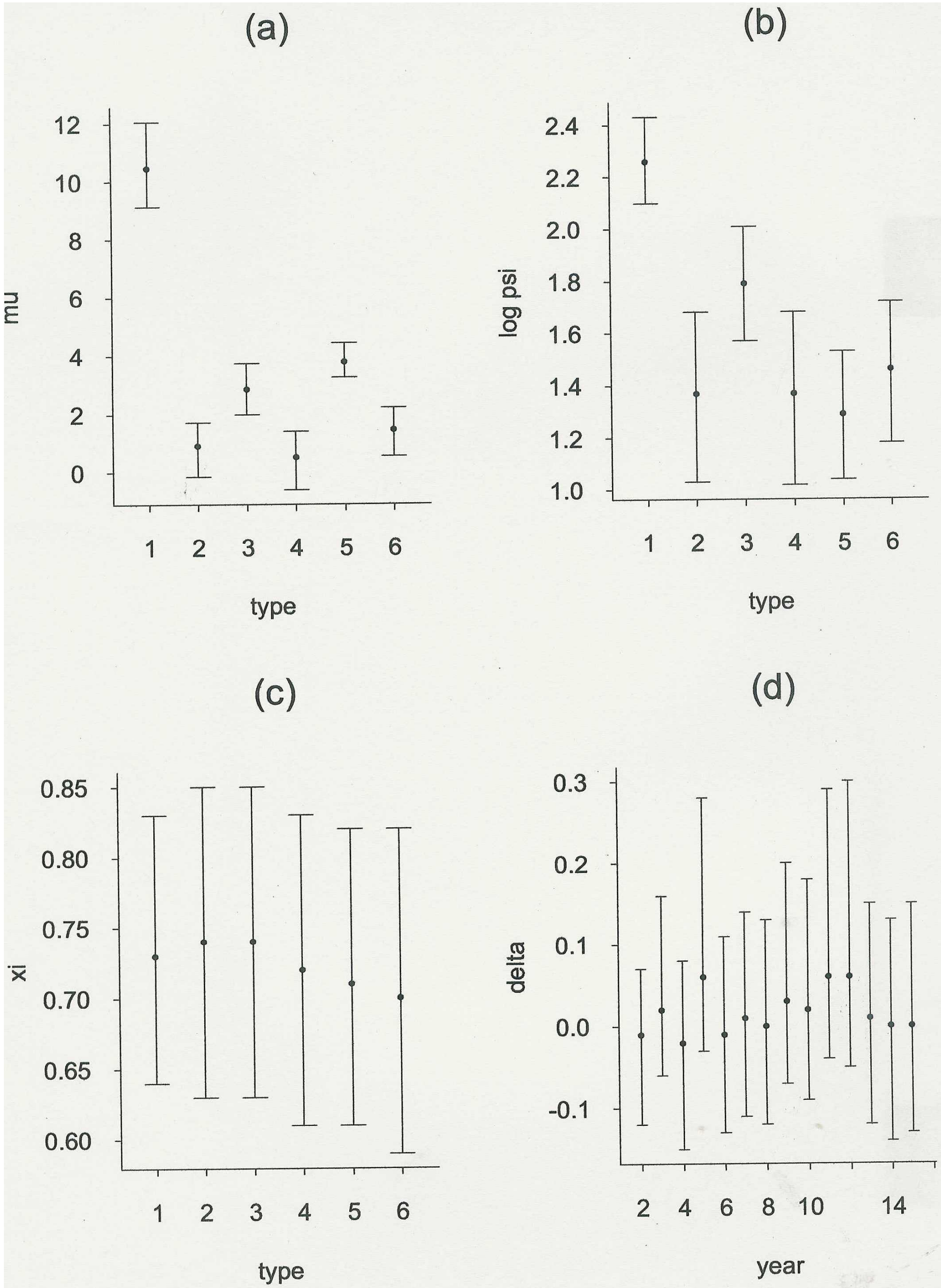
How can we use the fact that there are six different types of claim? One possibility is a *hierarchical model*, in which we assume each of the six types has separate extreme value parameters $\mu_j,\ \psi_j,\ \xi_j$, but that these have second-stage normal distributions,

$$\begin{aligned} \mu_j &\sim N[m_\mu, s_\mu^2],\ j = 1, ..., 6, \\ \log \psi_j &\sim N[m_\psi, s_\psi^2],\ j = 1, ..., 6, \\ \xi_j &\sim N[m_\xi, s_\xi^2],\ j = 1, ..., 6. \end{aligned}$$

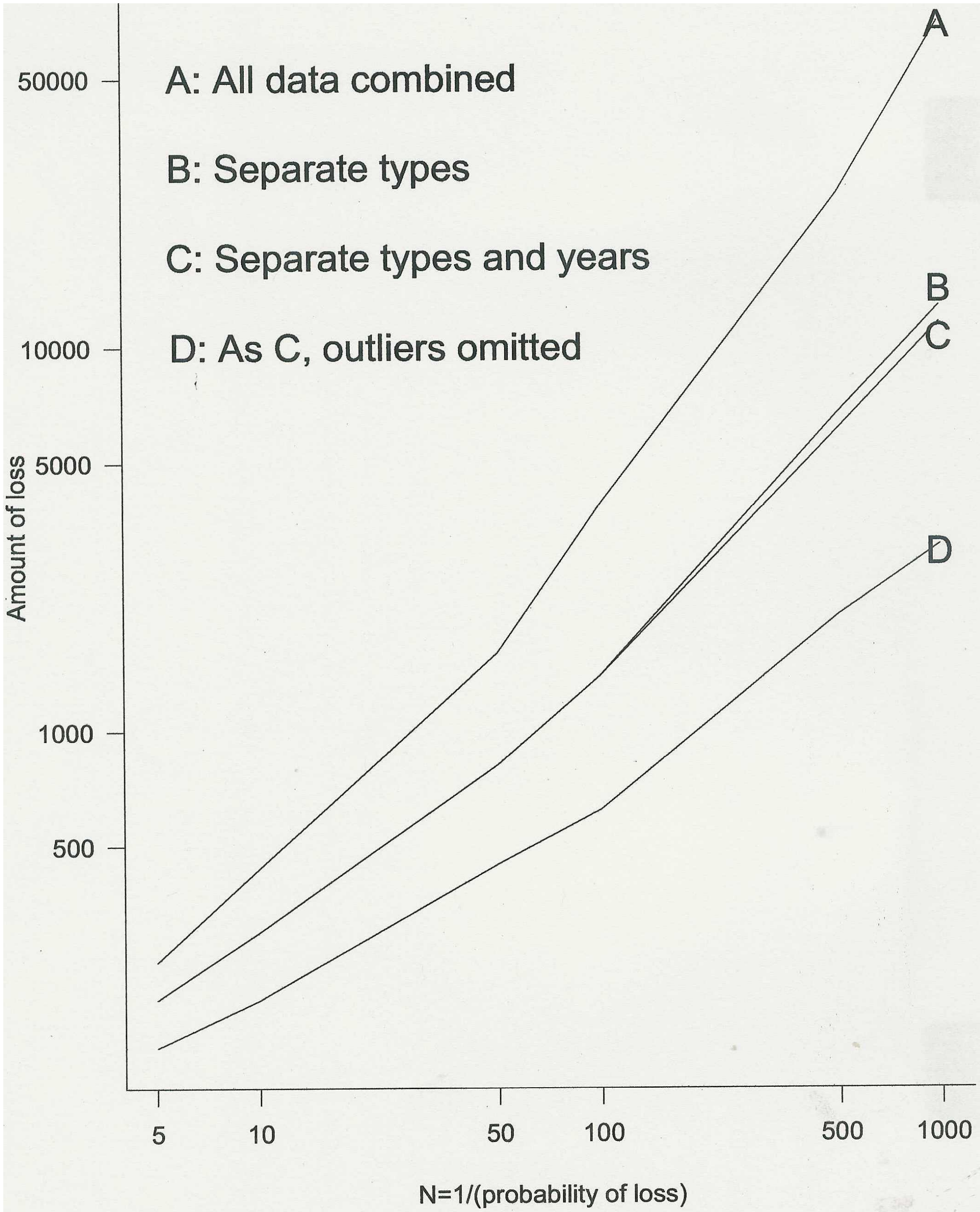
This can be fitted by hierarchical Bayesian methods.

In an extension of the same idea, we also assume that each year has a yearly parameter δ_k which is also drawn from a normal distribution.

We show boxplots for each of $\mu_j, \log \psi_j, \xi_j,\ j = 1, ..., 6$ and for $\delta_k,\ k = 2, 15$.



Posterior means and quartiles for μ_j, $\log \psi_j$, ξ_j $(j = 1, ..., 6)$ and for δ_k $(k = 2, ..., 15)$.



Computations of posterior predictive distribution functions (plotted on a log-log scale) corresponding to the homogenous model (curve A) and three different versions of the hierarchical model.

The final plot shows the advntage of the hierarchical approach. We actually get *less* extreme predictions under the hierarchical model (curves B or C) than we do ignoring the hierarchical structure (curve A).

IV. TREND IN PRECIPITATION EXTREMES

(joint work with Amy Grady and Gabi Hegerl)

During the past decade, there has been extensive research by climatologists documenting increases in the levels of extreme precipitation, but in observational and model-generated data.

With a few exceptions (papers by Katz, Zwiers and co-authors) this literature have not made use of the extreme value distributions and related constructs

There are however a few papers by statisticians that have explored the possibility of using more advanced extreme value methods (e.g. Cooley, Naveau and Nychka, to appear *JASA*; Sang and Gelfand, submitted)

This discussion uses extreme value methodology to look for trends

DATA SOURCES

- NCDC Rain Gauge Data (Groisman 2000)
  - Daily precipitation from 5873 stations
  - Select 1970–1999 as period of study
  - 90% data coverage provision — 4939 stations meet that

- NCAR-CCSM climate model runs
  - 20×41 grid cells of side 1.4^{o}
  - 1970–1999 and 2070–2099 (A2 scenario)

- PRISM data
  - 1405×621 grid, side 4km
  - Elevations
  - Mean annual precipitation 1970–1997

EXTREME VALUES METHODOLOGY

Based on "point process" extreme values methodology (cf. Smith 1989, Coles 2001, Smith 2003)

Inhomogeneous case:

- Time-dependent threshold u_t and parameters $\mu_t,\ \psi_t,\ \xi_t$

- Exceedance $y > u_t$ at time t has probability

$$\frac{1}{\psi_t}\left(1+\xi_t\frac{y-\mu_t}{\psi_t}\right)_+^{-1/\xi_t-1}\exp\left\{-\left(1+\xi_t\frac{u_t-\mu_t}{\psi_t}\right)_+^{-1/\xi_t}\right\}dydt$$

- Estimation by maximum likelihood

Seasonal models without trends

General structure:

$$
\begin{aligned}
\mu_t &= \theta_{1,1} + \sum_{k=1}^{K_1} \left(\theta_{1,2k} \cos \frac{2\pi kt}{365.25} + \theta_{1,2k+1} \sin \frac{2\pi kt}{365.25} \right), \\
\log \psi_t &= \theta_{2,1} + \sum_{k=1}^{K_2} \left(\theta_{2,2k} \cos \frac{2\pi kt}{365.25} + \theta_{2,2k+1} \sin \frac{2\pi kt}{365.25} \right), \\
\xi_t &= \theta_{3,1} + \sum_{k=1}^{K_3} \left(\theta_{3,2k} \cos \frac{2\pi kt}{365.25} + \theta_{3,2k+1} \sin \frac{2\pi kt}{365.25} \right).
\end{aligned}
$$

Call this the (K_1, K_2, K_3) model.

Note: This is all for one station. The θ parameters will differ at each station.

Models with trend

Add to the above:

- Overall linear trend $\theta_{j,2K+2}t$ added to any of μ_t ($j = 1$), $\log \psi_t$ ($j = 1$), ξ_t ($j = 1$). Define K_j^* to be 1 if this term is included, o.w. 0.

- Interaction terms of form

$$t \cos \frac{2\pi kt}{365.25}, \quad t \sin \frac{2\pi kt}{365.25}, \quad k = 1, ..., K_j^{**}.$$

Typical model denoted

$$(K_1, K_2, K_3) \times (K_1^*, K_2^*, K_3^*) \times (K_1^{**}, K_2^{**}, K_3^{**})$$

Eventually use $(4, 2, 1) \times (1, 1, 0) \times (2, 2, 0)$ model (27 parameters for each station)

SPATIAL SMOOTHING

Let Z_s be field of interest, indexed by s (typically the logarithm of the 25-year RV at site s, or a log of ratio of RVs. Taking logs improves fit of spatial model, to follow.)

Don't observe Z_s — estimate $\hat{Z}_s$. Assume

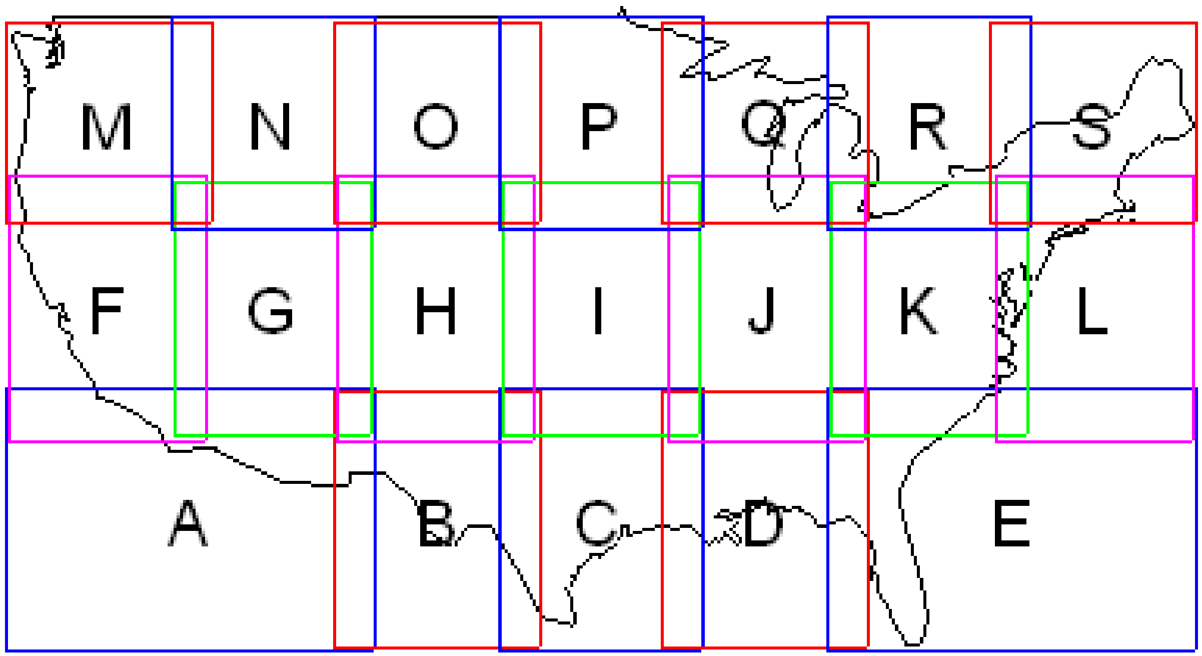
$$\begin{aligned} \hat{Z} \mid Z &\sim N[Z, W] \\ Z &\sim N[X\beta, V(\phi)] \\ \hat{Z} &\sim N[X\beta, V(\phi) + W]. \end{aligned}$$

for known W; X are covariates, β are unknown regression parameters and ϕ are parameters of spatial covariance matrix V.

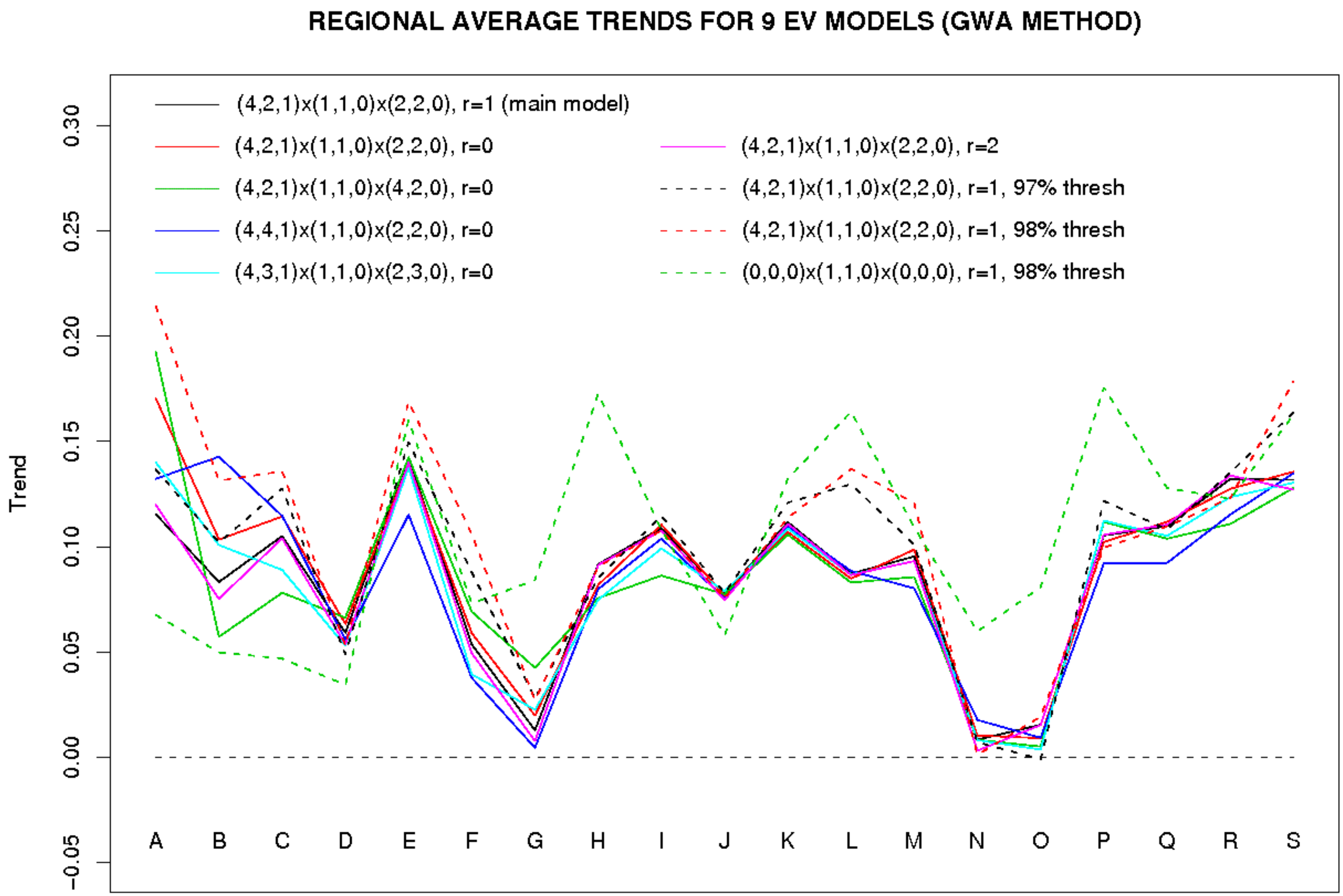
- ϕ by REML
- β given ϕ by GLS
- Predict Z at observed and unobserved sites by kriging

Spatial Heterogeneity

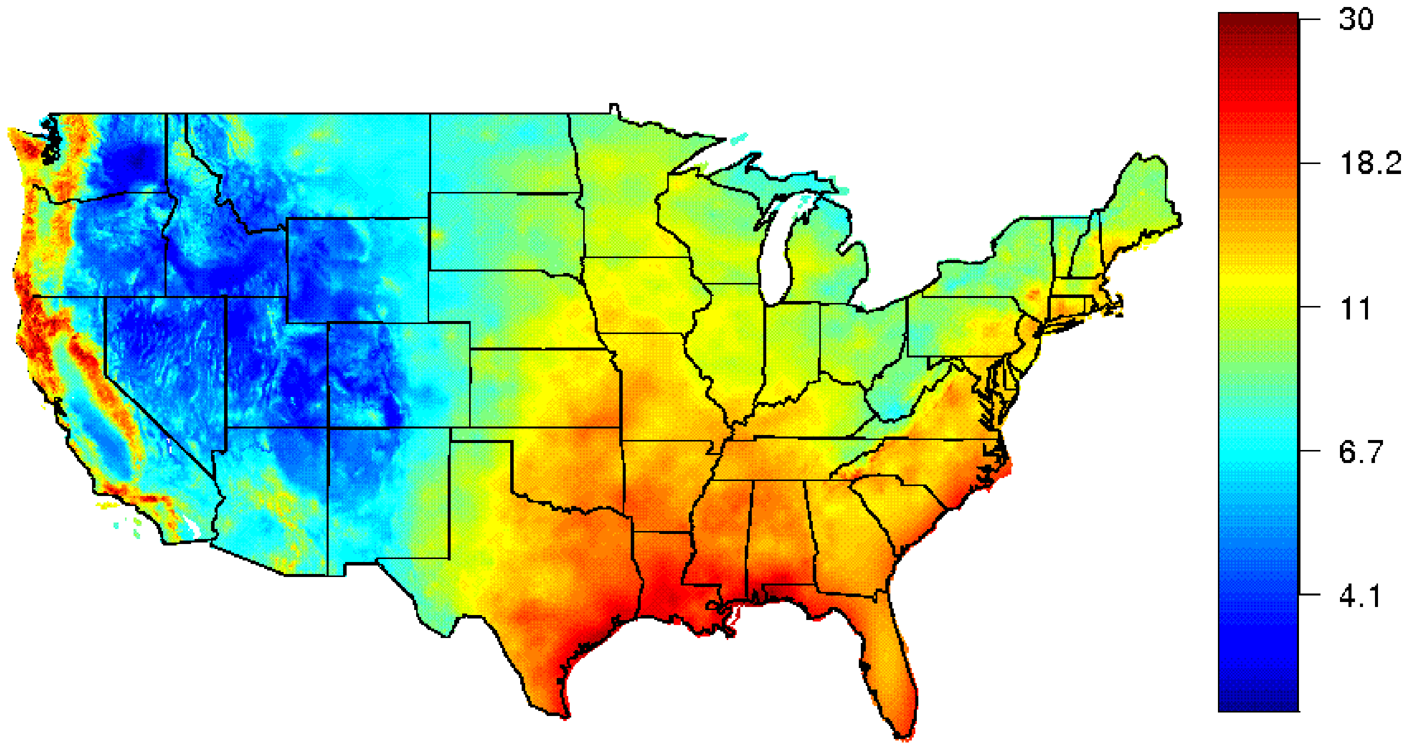
- Divide US into 19 overlapping regions, most $10^{o} \times 10^{o}$
  - Kriging within each region
  - Linear smoothing across region boundaries
  - Same for MSPEs
  - Also calculate regional averages, including MSPE



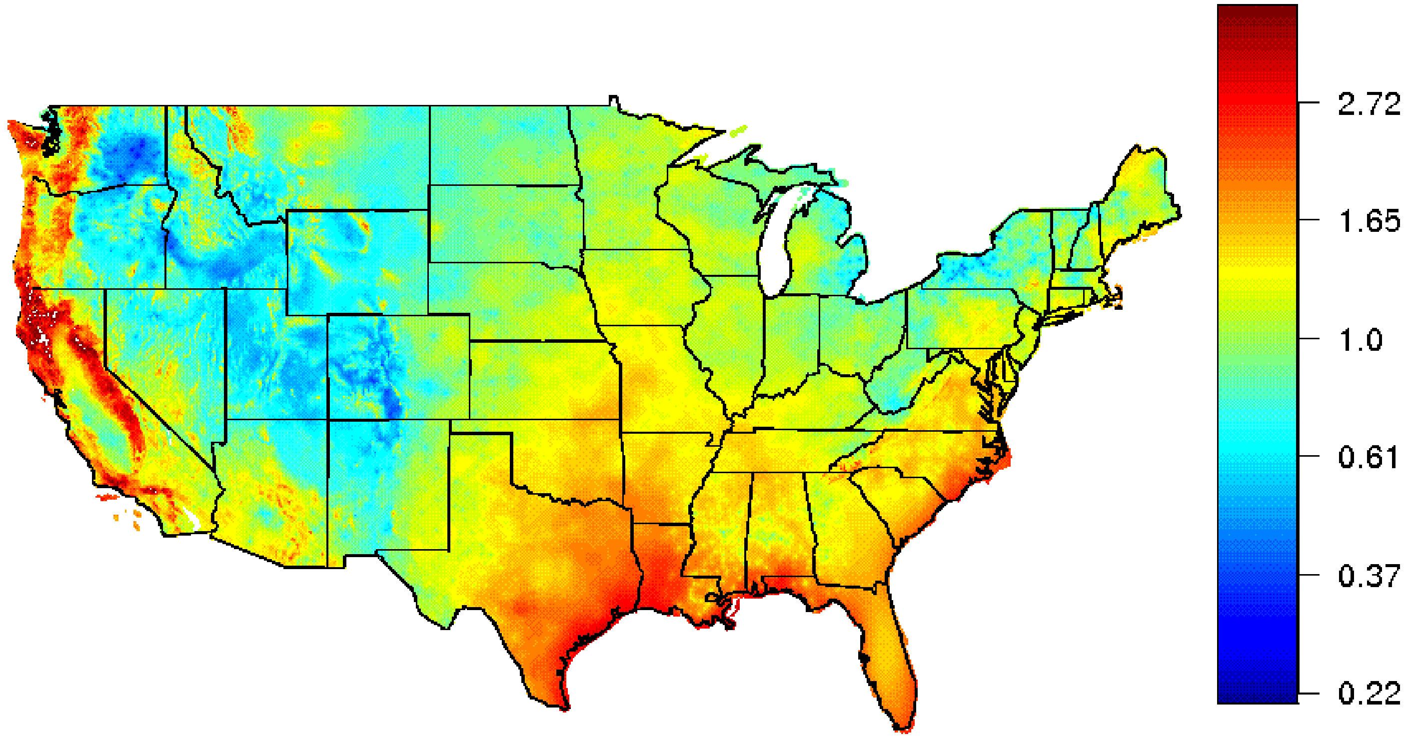
Continental USA divided into 19 regions



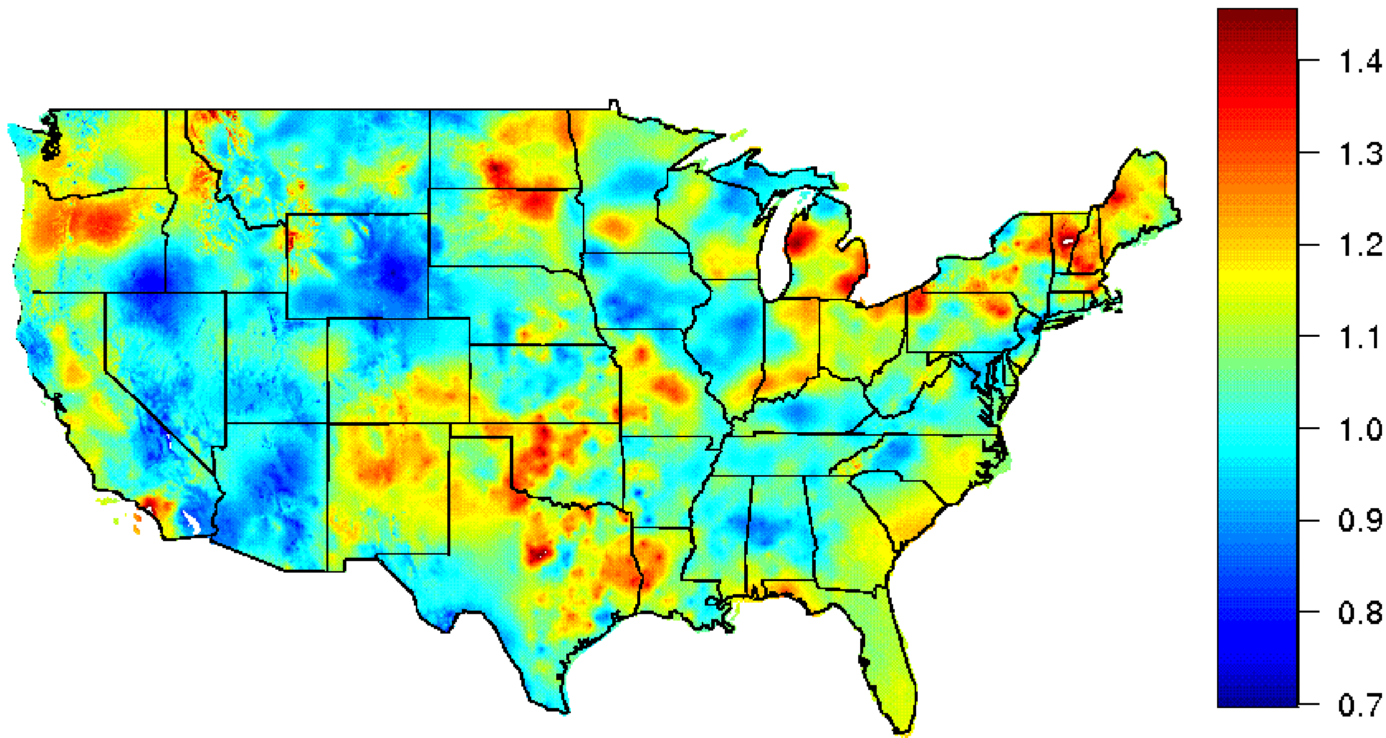
Trends across 19 regions (measured as change in log RV25) for 8 different seasonal models and one non-seasonal model with simple linear trends. Regional averaged trends by geometric weighted average approach.



Map of 25-year return values (cm.) for the years 1970–1999



Root mean square prediction errors for map of 25-year return values for 1970–1999



Ratios of return values in 1999 to those in 1970

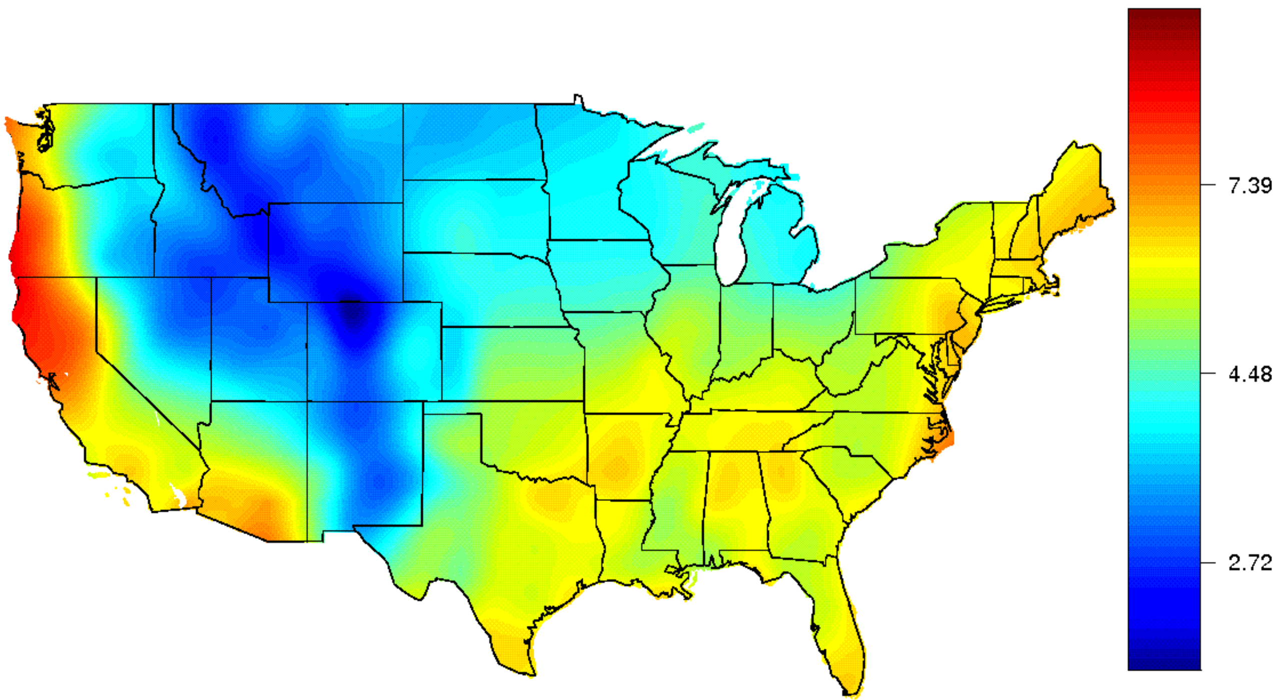
	Δ_1	S_1	Δ_2	S_2		Δ_1	S_1	Δ_2	S_2
A	−0.01	.03	0.05^{**}	.05	K	0.08^{***}	.01	0.09^{**}	.03
B	0.07^{**}	.03	0.08^{***}	.04	L	0.07^{***}	.02	0.07^{*}	.04
C	0.11^{***}	.01	0.10	.03	M	0.07^{***}	.02	0.10^{**}	.03
D	0.05^{***}	.01	0.06	.05	N	0.02	.03	0.01	.03
E	0.13^{***}	.02	0.14^{*}	.05	O	0.01	.02	0.02	.03
F	0.00	.02	0.05^{*}	.04	P	0.07^{***}	.01	0.11^{***}	.03
G	−0.01	.02	0.01	.03	Q	0.07^{***}	.01	0.11^{***}	.03
H	0.08^{***}	.01	0.10^{***}	.03	R	0.15^{***}	.02	0.13^{***}	.03
I	0.07^{***}	.01	0.12^{***}	.03	S	0.14^{***}	.02	0.12^{*}	.06
J	0.05^{***}	.01	0.08^{**}	.03					

Δ_1: Mean change in log 25-year return value (1970 to 1999) by kriging

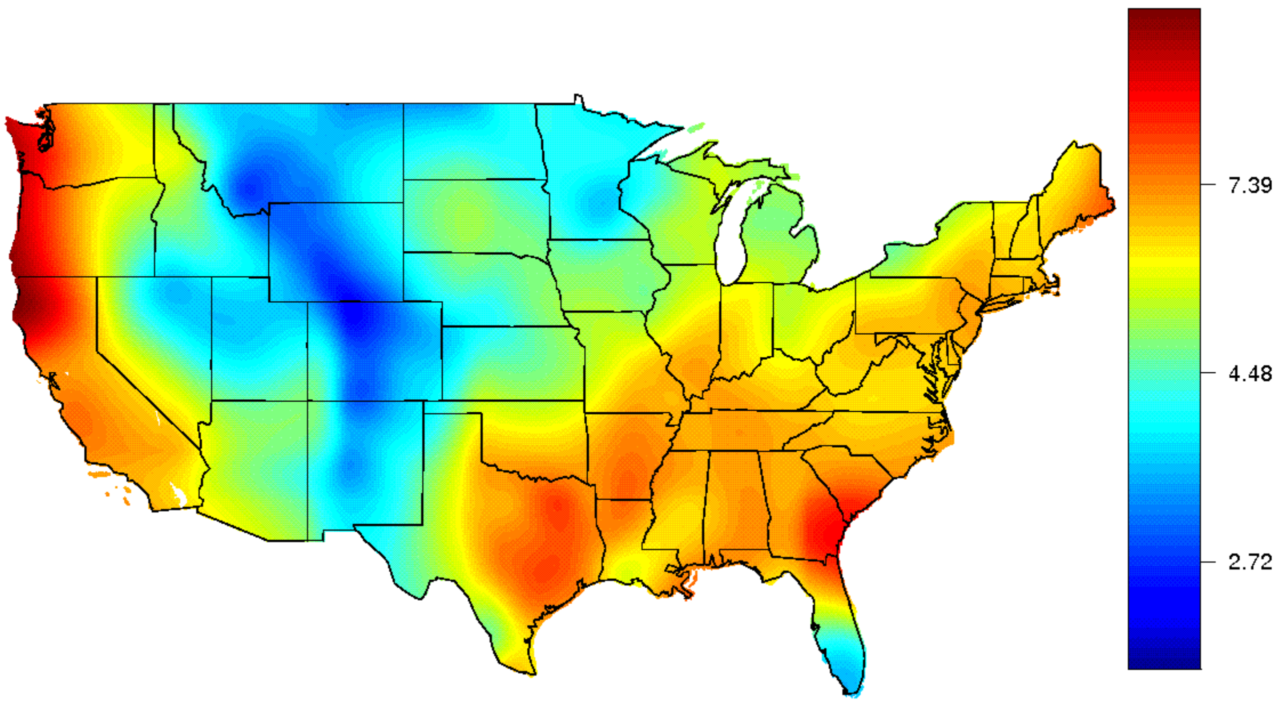
S_1: Corresponding standard error (or RMSPE)

Δ_2, S_2: same but using geometrically weighted average (GWA)

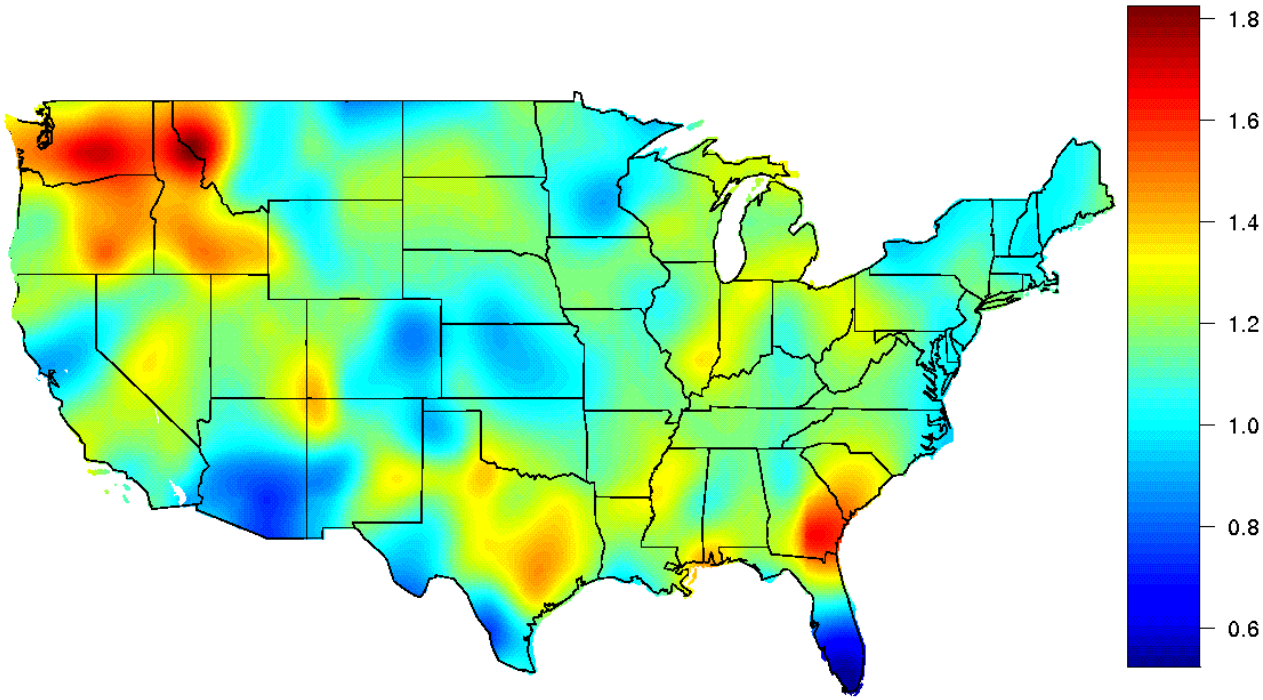
Stars indicate significance at 5%*, 1%**, 0.1%***.



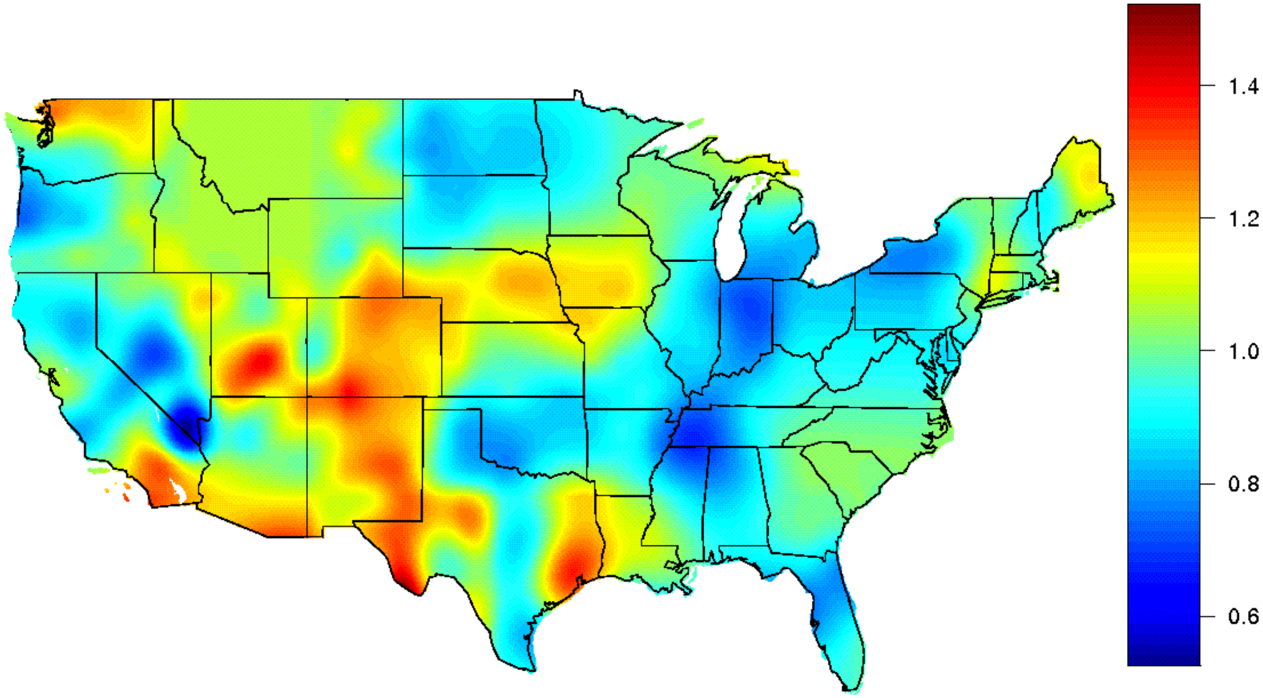
Return value map for CCSM data (cm.): 1970–1999



Return value map for CCSM data (cm.): 2070–2099



Estimated ratios of 25-year return values for 2070–2099 to those of 1970–1999, based on CCSM data, A2 scenario



Extreme value model with trend: ratio of 25-year return value in 1999 to 25-year return value in 1970, based on CCSM data

CONCLUSIONS

1. Focus on N-year return values — strong historical tradition for this measure of extremes (we took $N = 25$ here)

2. Seasonal variation of extreme value parameters is a critical feature of this analysis

3. Overall significant increase over 1970–1999 except for parts of western states — average increase across continental US is 7%

4. Projections to 2070–2099 show further strong increases but note caveat based on point 5

5. *But...* based on CCSM data there is a completely different spatial pattern and no overall increase — still leaves some doubt as to overall interpretation.

FURTHER READING

Finkenstadt, B. and Rootzén, H. (editors) (2003), *Extreme Values in Finance, Telecommunications and the Environment.* Chapman and Hall/CRC Press, London.
(See http://www.stat.unc.edu/postscript/rs/semstatrls.pdf)

Coles, S.G. (2001), *An Introduction to Statistical Modeling of Extreme Values.* Springer Verlag, New York.

THANK YOU FOR YOUR ATTENTION!