HIERARCHICAL MODELS IN EXTREME VALUE THEORY

Richard L. Smith
Department of Statistics and Operations Research,
University of North Carolina, Chapel Hill
and

Statistical and Applied
Mathematical Sciences
Institute

Department of Statistics
North Carolina State University
February 19, 2015

OUTLINE OF TALKs

I. An Example of Hierarchical Models Applied to Insurance Extremes

II. Attribution of Climate Extremes

III. The Attribution Problem for Joint Distributions of Climate Extremes (introduction)

I. AN EXAMPLE OF HIERARCHICAL MODELS APPLIED TO INSURANCE EXTREMES

From the book chapter *Bayesian Risk Analysis* by R.L. Smith and D.J. Goodman (2000)

http://www.stat.unc.edu/postscript/rs/pred/inex1.pdf

See also:

R.L. Smith (2003), Statistics of Extremes, With Applications in Environment, Insurance and Finance. In *Extreme Values in Finance, Telecommunications and the Environment*, edited by B. Finkenstadt and H. Rootzen, Chapman and Hall/CRC Press, London, pp. 1-78.

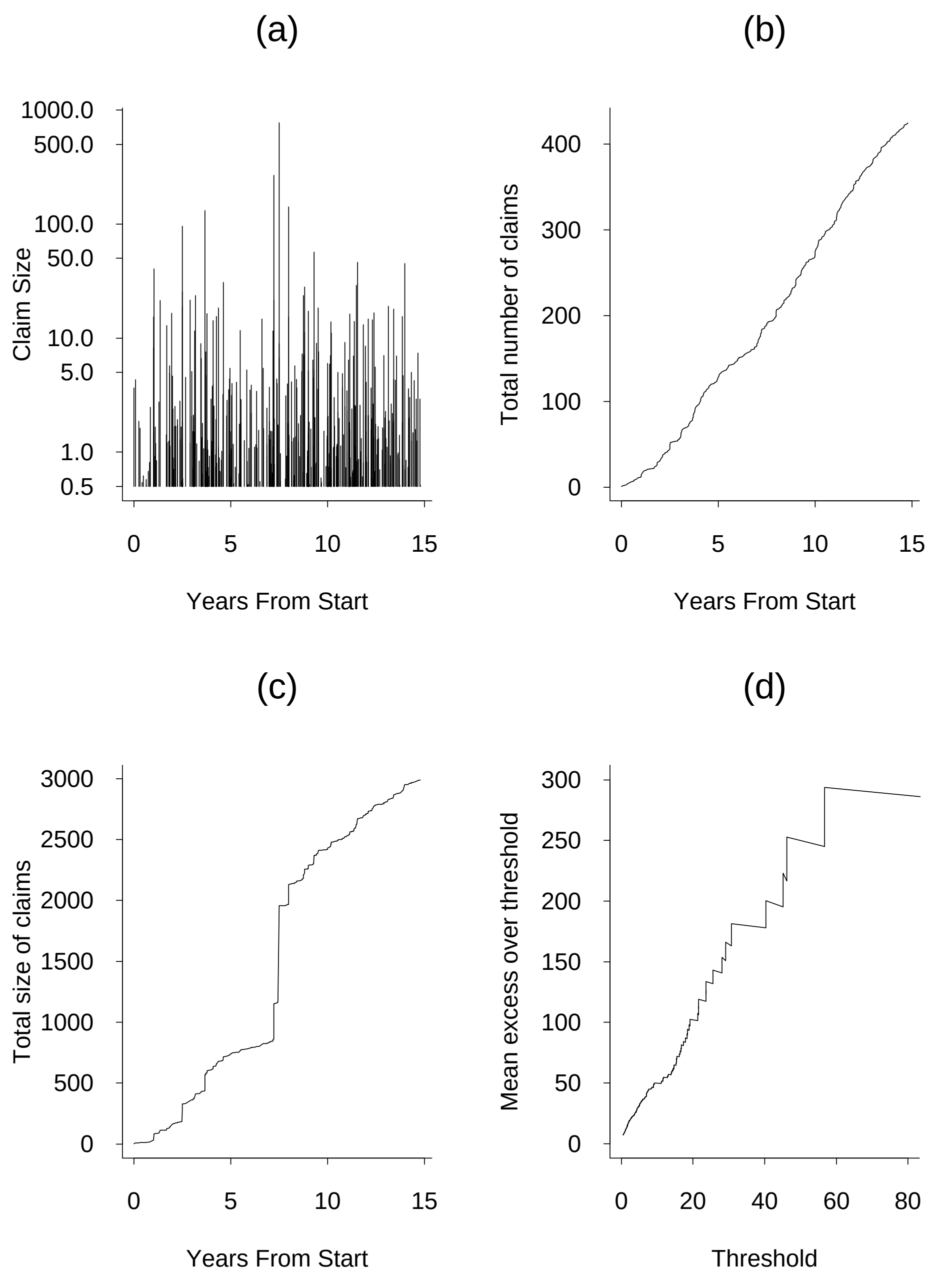
http://www.stat.unc.edu/postscript/rs/semstatrls.pdf

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?

Introduction to Univariate Extreme Value Theory

EXTREME VALUE DISTRIBUTIONS

$X_1, X_2, \ldots$, i.i.d., $F(x) = \Pr\{X_i \le x\}$, $M_n = \max(X_1, \ldots, X_n)$,

$$\Pr\{M_n \le x\} = F(x)^n.$$

For non-trivial results must *renormalize*: find $a_n > 0, b_n$ such that

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

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)

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

Equivalence to three types theorem established by Pickands (1975).

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 $-\sigma/\xi$.

POISSON-GPD MODEL FOR EXCEEDANCES

1. The number, N, of exceedances of the level u in any one year has a Poisson distribution with mean λ,

2. Conditionally on $N \geq 1$, the excess values $Y_1, ..., Y_N$ are **IID** from the GPD.

Relation to GEV for annual maxima:

Suppose $x > u$. The probability that the annual maximum of the Poisson-GPD process is less than x is

$$
\begin{aligned}
& \mathrm{Pr}\{\max_{1 \le i \le N} Y_i \le x\} = \mathrm{Pr}\{N = 0\} + \sum_{n=1}^{\infty} \mathrm{Pr}\{N = n,\ Y_1 \le x,\ \ldots\ Y_n \le x\} \\
= \ & e^{-\lambda} + \sum_{n=1}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} \left\{ 1 - \left(1 + \xi \frac{x-u}{\sigma} \right)^{-1/\xi} \right\}^n \\
= \ & \exp \left\{ -\lambda \left(1 + \xi \frac{x-u}{\sigma} \right)^{-1/\xi} \right\}.
\end{aligned}
$$

This is GEV with $\sigma = \psi + \xi(u - \mu), \lambda = \left(1 + \xi \frac{u-\mu}{\psi} \right)^{-1/\xi}$. Thus the GEV and GPD models are entirely consistent with one another above the GPD threshold, and moreover, shows exactly how the Poisson–GPD parameters σ and λ vary with u.

ALTERNATIVE PROBABILITY MODELS

1. The r largest order statistics model

If $Y_{n,1} \geq Y_{n,2} \geq ... \geq Y_{n,r}$ are r largest order statistics of IID sample of size n, and a_n and b_n are EVT normalizing constants, then

$$\left(\frac{Y_{n,1}-b_n}{a_n},...,\frac{Y_{n,r}-b_n}{a_n}\right)$$

converges in distribution to a limiting random vector $(X_1,...,X_r)$, whose density is

$$\begin{aligned} h(x_1,...,x_r) &= \psi^{-r}\exp\left\{-\left(1+\xi\frac{x_r-\mu}{\psi}\right)^{-1/\xi}\right. \\ &\quad \left. -\left(1+\frac{1}{\xi}\right)\sum_{j=1}^{r}\log\left(1+\xi\frac{x_j-\mu}{\psi}\right)\right\}. \end{aligned}$$

2. Point process approach (Smith 1989)

Two-dimensional plot of exceedance times and exceedance levels forms a nonhomogeneous Poisson process with

$$\begin{aligned} \Lambda(A) &= (t_2 - t_1)\Psi(y;\mu,\psi,\xi) \\ \Psi(y;\mu,\psi,\xi) &= \left(1 + \xi \frac{y-\mu}{\psi}\right)^{-1/\xi} \end{aligned}$$

$(1 + \xi(y-\mu)/\psi > 0)$.

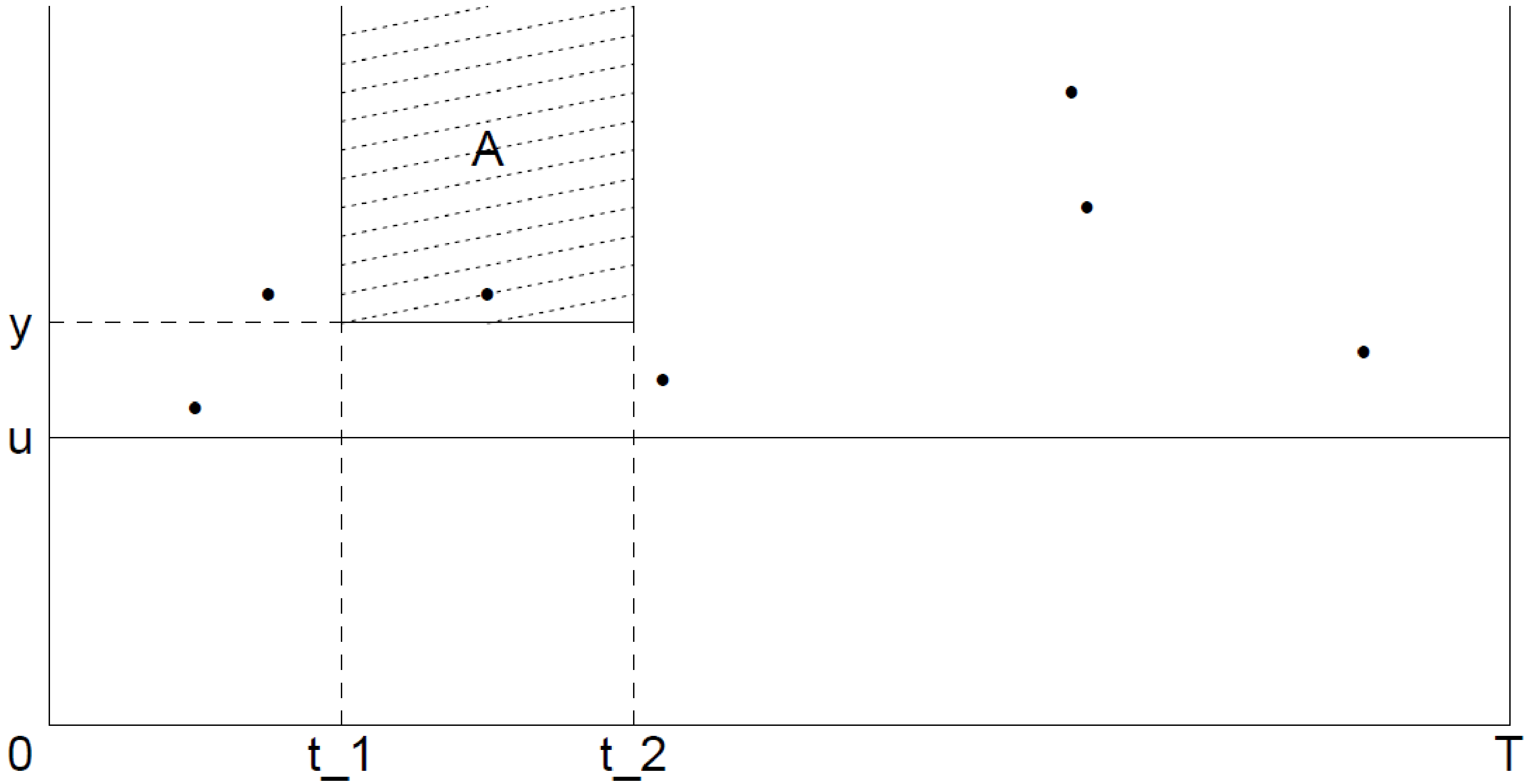


Illustration of point process model.

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

Comment. The point process approach is *almost* equivalent to the following: assume the GEV (not GPD) distribution is valid for exceedances over the threshold, and that all observations under the threshold are censored. Compared with the GPD approach, the parameterization directly in terms of μ, ψ, ξ is often easier to interpret, especially when trends are involved.

ESTIMATION

GEV log likelihood:

$$\begin{aligned}\ell_Y(\mu,\psi,\xi) &= -N\log\psi - \left(\frac{1}{\xi}+1\right)\sum_i \log\left(1+\xi\frac{Y_i-\mu}{\psi}\right)\\ &\quad -\sum_i\left(1+\xi\frac{Y_i-\mu}{\psi}\right)^{-1/\xi}\end{aligned}$$

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

Poisson-GPD model:

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

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

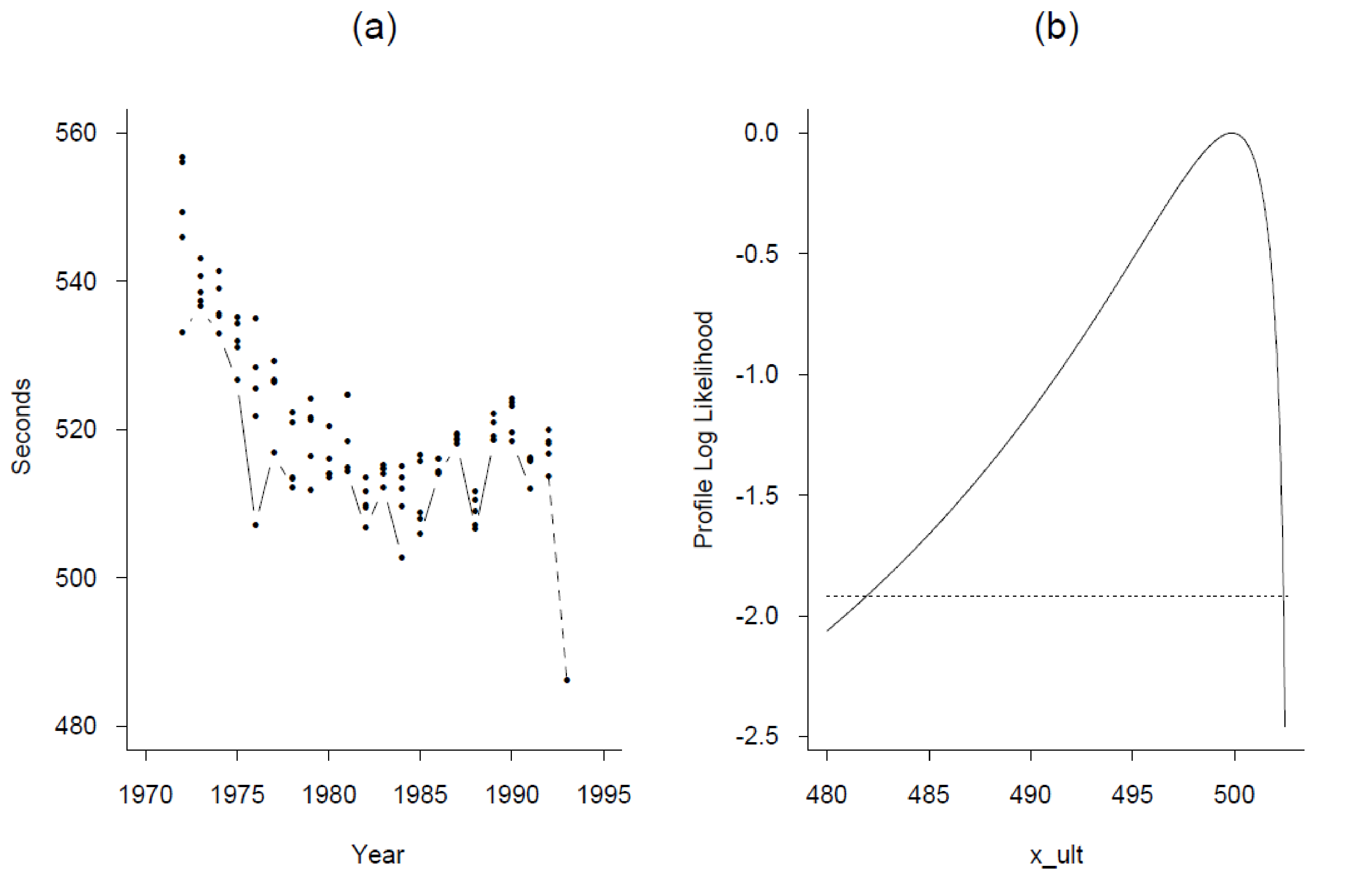
Usual asymptotics valid if $\xi>-\frac{1}{2}$ (Smith 1985)

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 $\mathrm{Pr}\{Z > z\}$ by

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

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



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.

Insurance Extremes Dataset

We return to the oil company data set discussed earlier. Prior to any of the analysis, some examination was made of clustering phenomena, but this only reduced the original 425 claims to 393 "independent" claims (Smith & Goodman 2000)

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:

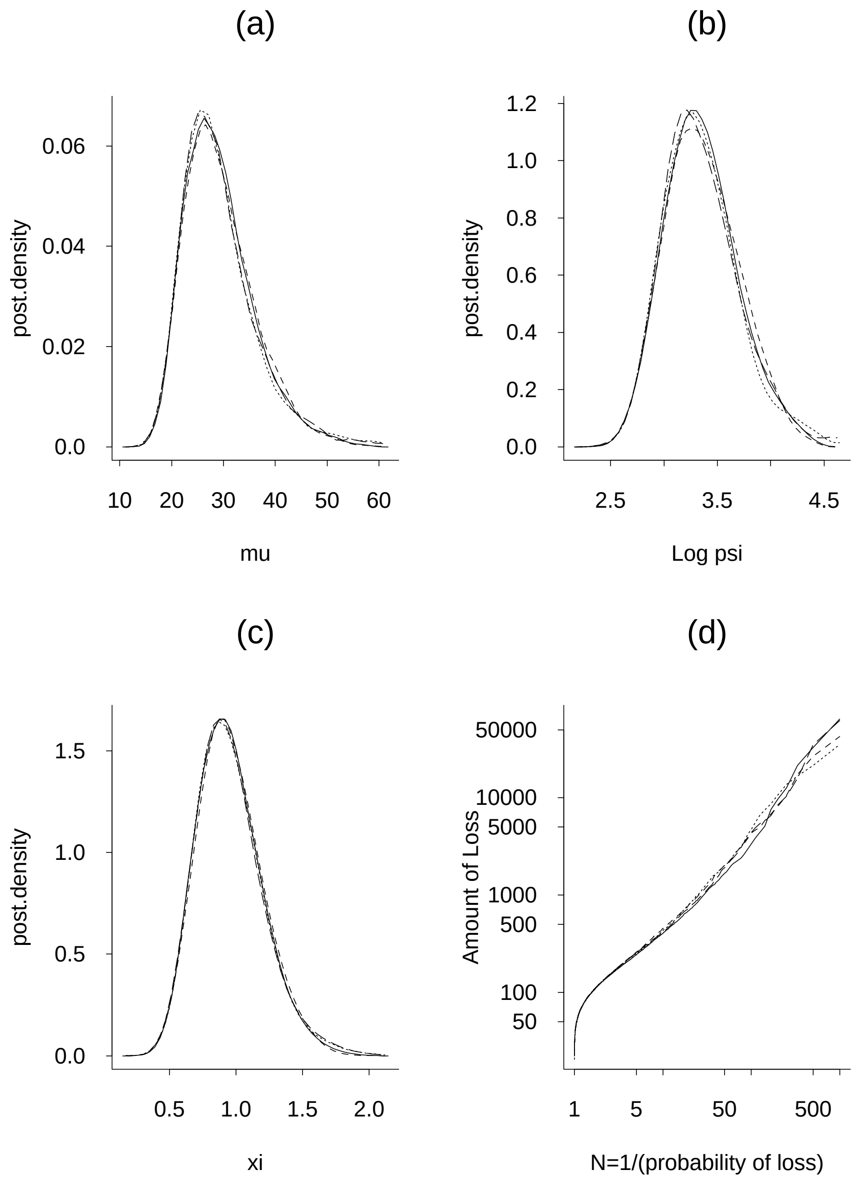
$$\widehat{G}(y) = G(y; \widehat{\mu}, \widehat{\psi}, \widehat{\xi})$$

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

Further features of the data:

1. When separate GPDs are fitted to each of the 6 main types, there are clear differences among the parameters.

2. The rate of high-threshold crossings does not appear uniform, but peaks around years 10–12.

A Hierarchical Model:

Level I. Parameters $m_\mu,\ m_\psi,\ m_\xi, s_\mu^2,\ s_\psi^2,\ s_\xi^2$ are generated from a prior distribution.

Level II. Conditional on the parameters in Level I, parameters $\mu_1, ..., \mu_J$ (where J is the number of types) are independently drawn from $N(m_\mu, s_\mu^2)$, the normal distribution with mean m_μ, variance s_μ^2. Similarly, $\log\psi_1, ..., \log\psi_J$ are drawn independently from $N(m_\psi, s_\psi^2)$, $\xi_1, ..., \xi_J$ are drawn independently from $N(m_\xi, s_\xi^2)$.

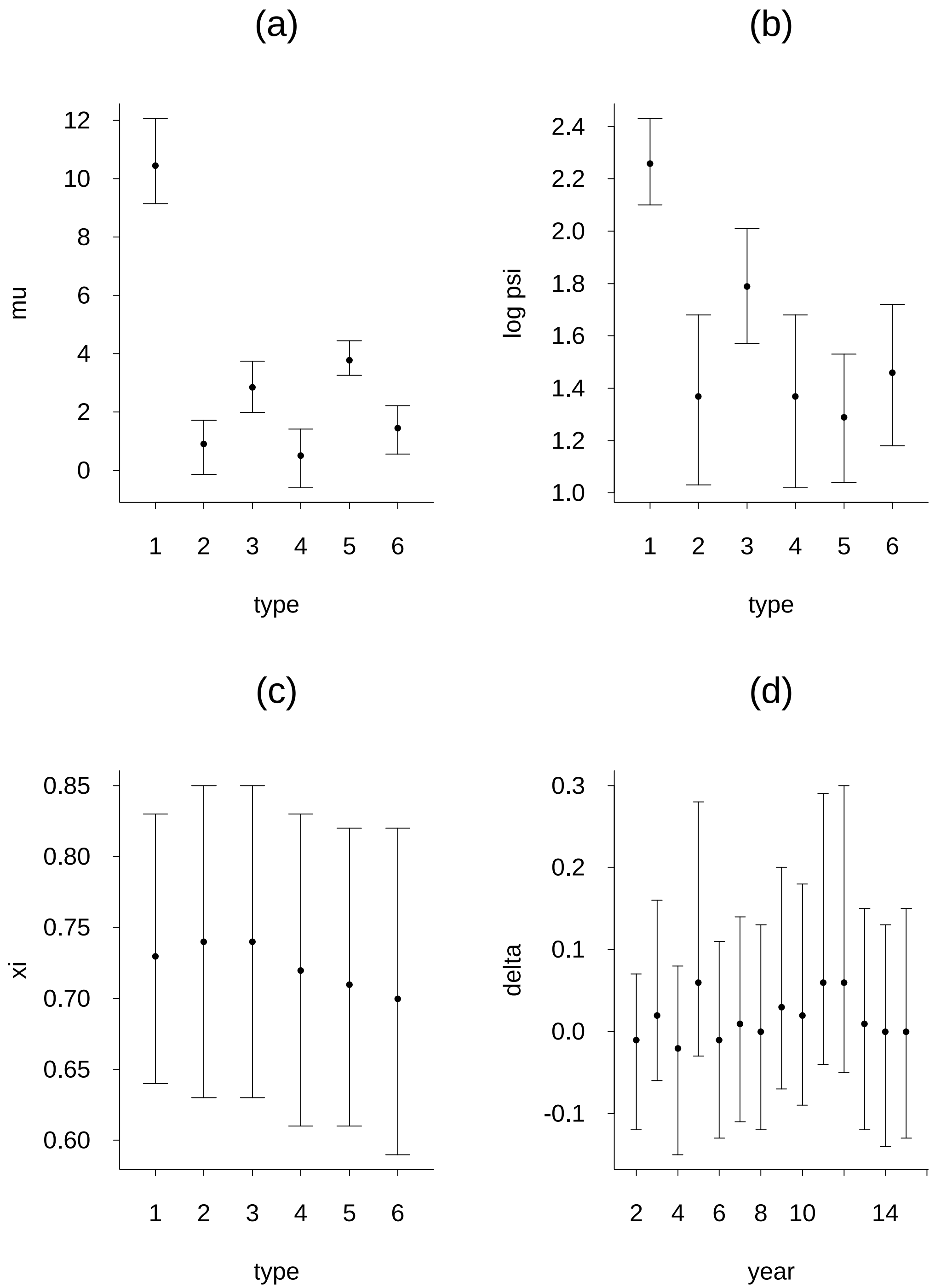
Level III. Conditional on Level II, for each $j \in \{1, ..., J\}$, the point process of exceedances of type j is generated from the Poisson process with parameters $\mu_j,\ \psi_j,\ \xi_j$.

This model may be further extended to include a year effect, as follows. Suppose the extreme value parameters for type j in year k are not μ_j, ψ_j, ξ_j but $\mu_j + \delta_k, \psi_j, \xi_j$. We fix $\delta_1 = 0$ to ensure identifiability, and let $\{\delta_k,\ k > 1\}$ follow an AR(1) process:

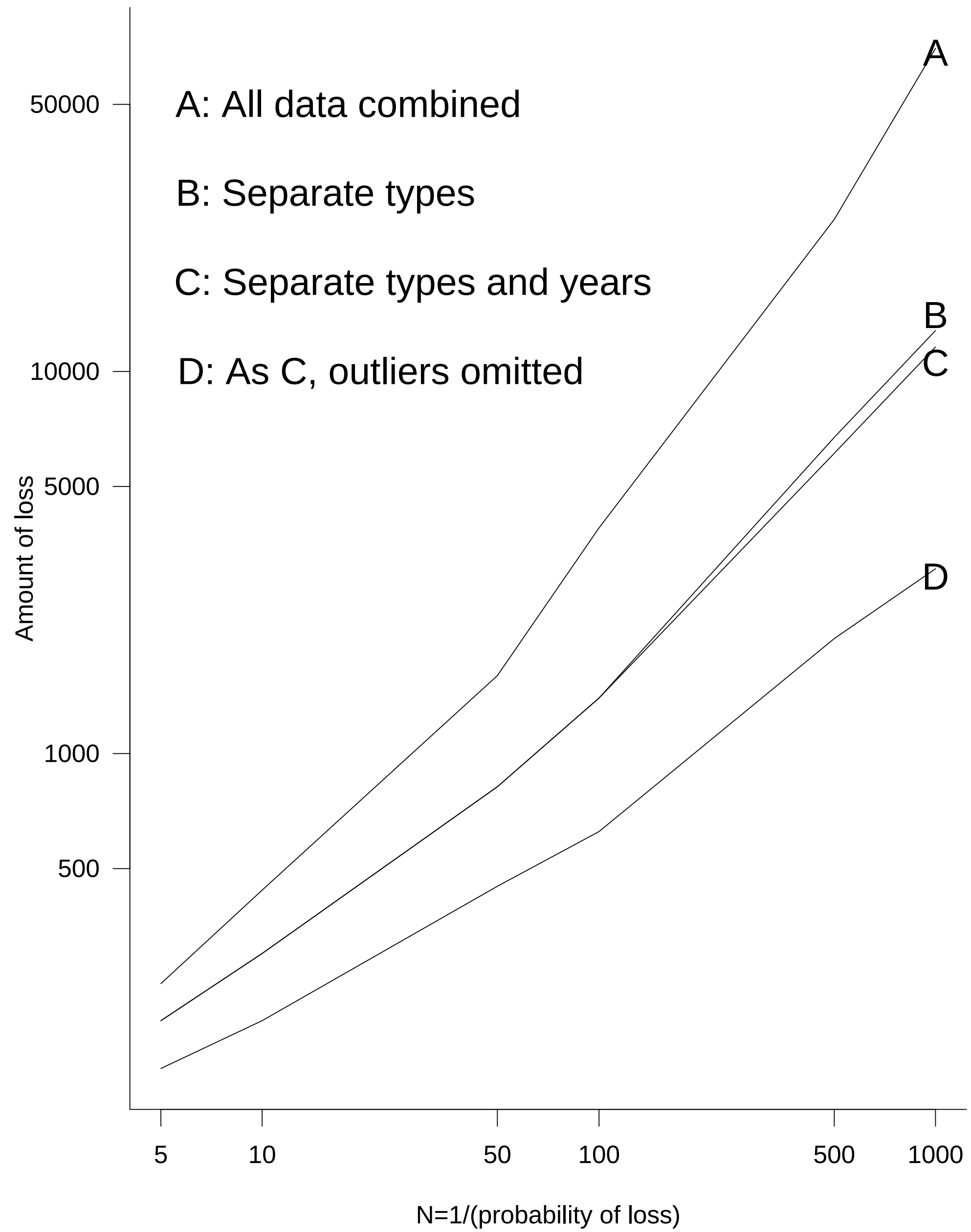
$$\delta_k = \rho\delta_{k-1} + \eta_k, \quad \eta_k \sim N(0, s_\eta^2)$$

with a vague prior on (ρ, s_η^2).

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



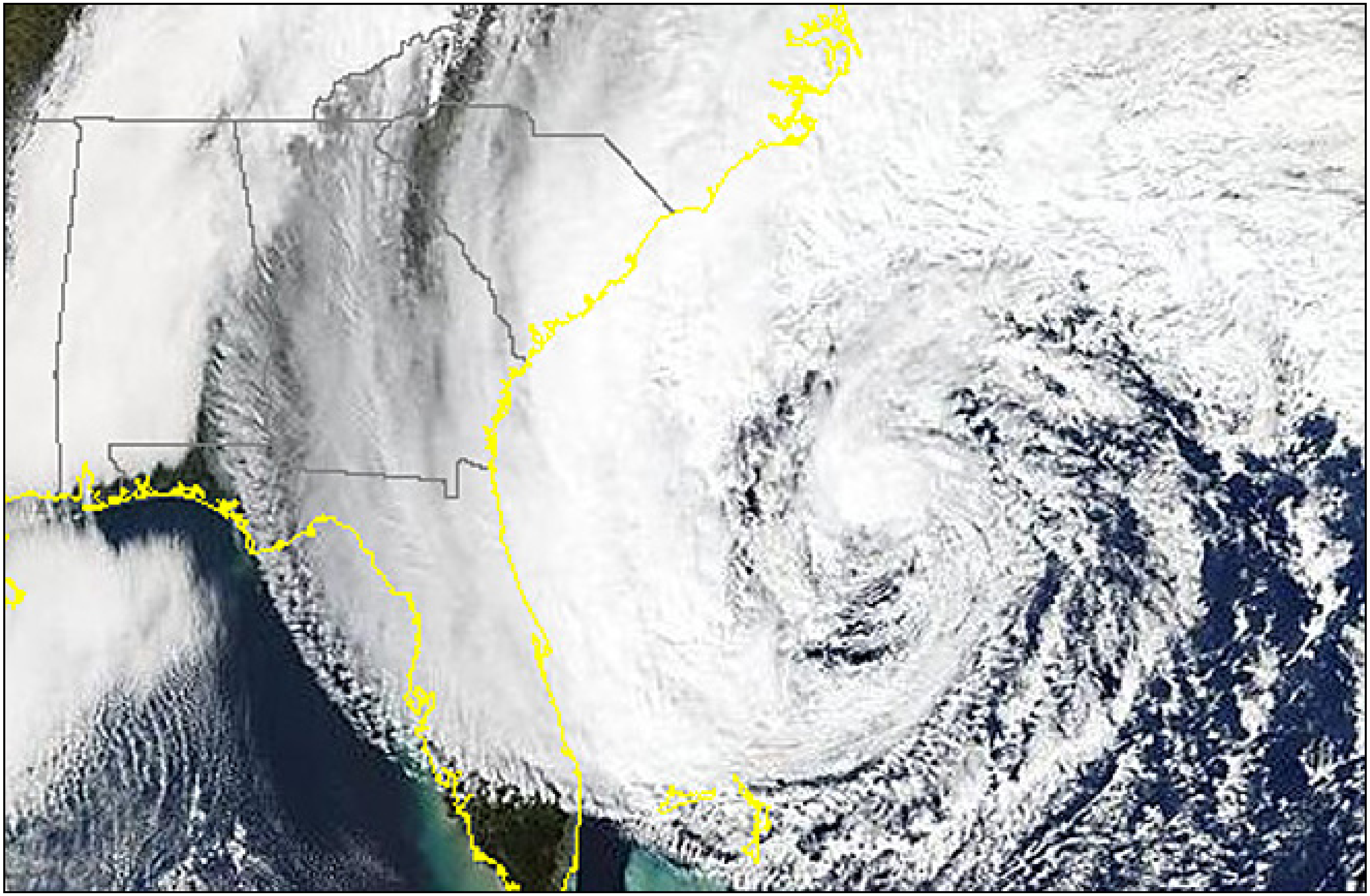
Posterior means, quartiles for μ_j, $\log\psi_j$, ξ_j $(j = 1, \ldots, 6)$ and δ_k $(k = 2, \ldots, 15)$.



Posterior predictive distribution functions (log-log scale) for homogenous model (curve A) and three versions of hierarchical model

II. ATTRIBUTION OF CLIMATE EXTREMES

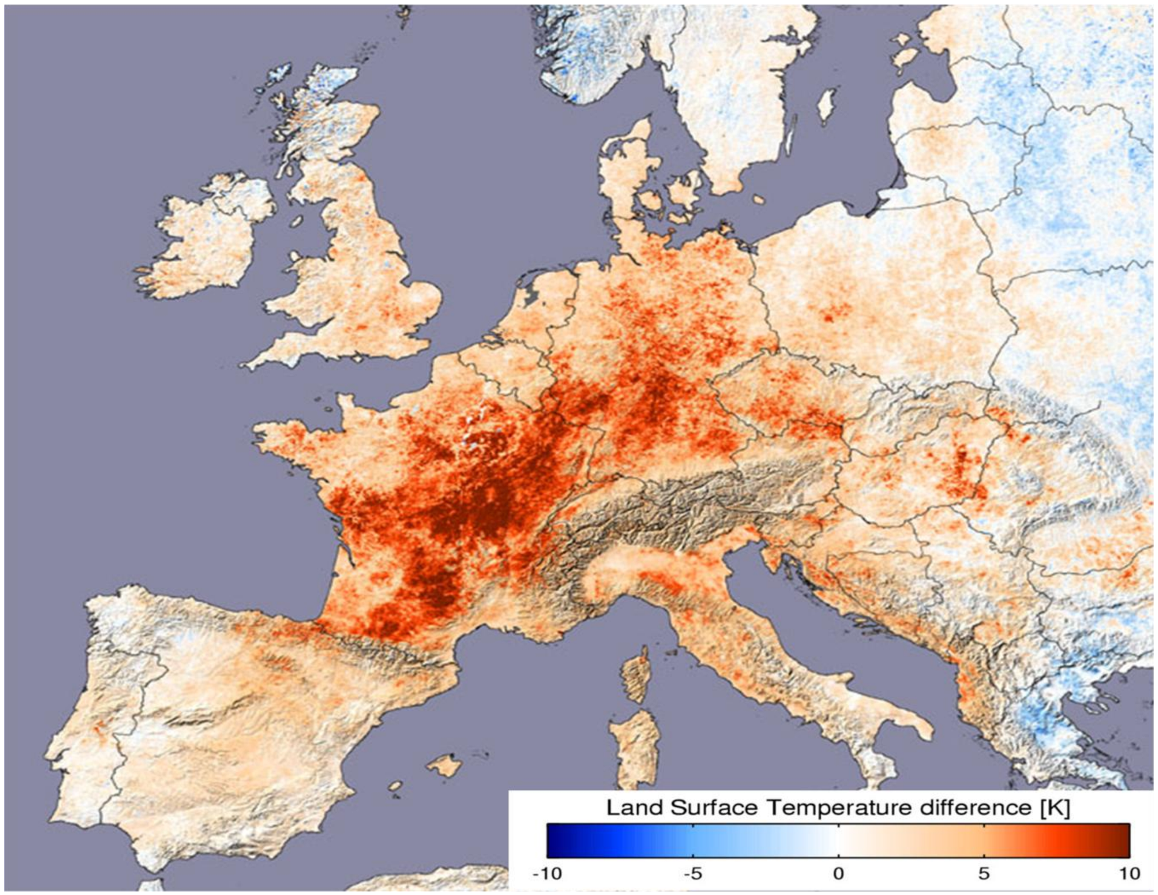
(Joint work with Michael Wehner, Lawrence Berkeley National Laboratory)



Superstorm Sandy on October 27 2012 (Scott Sistek)

Superstorm Sandy (www.guardian.co.uk; October 30, 2012)

Superstorm Sandy (www.cnn.com; October 31, 2012)



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

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

Motivating Question:

- Concern over increasing frequency of extreme meteorological events
  - Is the increasing frequency a result of anthropogenic influence?
  - How much more rapidly with they increase in the future?

- Focus on three specific events: heatwaves in Europe 2003, Russia 2010 and Central USA 2011

- Identify meteorological variables of interest — JJA temperature averages over a region
  - Europe — 10^o W to 40^o E, 30^o to 50^o N
  - Russia — 30^o to 60^o E, 45^o to 65^o N
  - Central USA — 90^o to 105^o W, 25^o to 45^o N

- Probabilities of crossing thresholds — respectively 1.92K, 3.65K, 2.01K — in any year from 1990 to 2040.

Background

- Stott, Stone and Allen (2004) defined "fraction of attributable risk" (FAR)
  - Observe some extreme event
  - Let P_1 be probability of this event estimated from models including anthropogenic forcings, P_0 corresponding probability under natural forcings
  - $FAR = 1 - \frac{P_0}{P_1}$ (also consider $RR = \frac{P_1}{P_0}$)

- For the Europe 2003 event they claimed $P_1 \approx \frac{1}{250}$, $P_0 \approx \frac{1}{1000}$ so $RR = 4$ and $FAR = 1 - \frac{1}{4} = 0.75$.

- "Very likely" (confidence level at least 90%) that the FAR was at least 0.5 ($RR \geq 2$),

- Method used a combination of extreme value theory, and detection and attribution methodology from the climate literature