EXTREME VALUE THEORY

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

Notes for STOR 890 Course
Chapel Hill, Spring Semester 2009

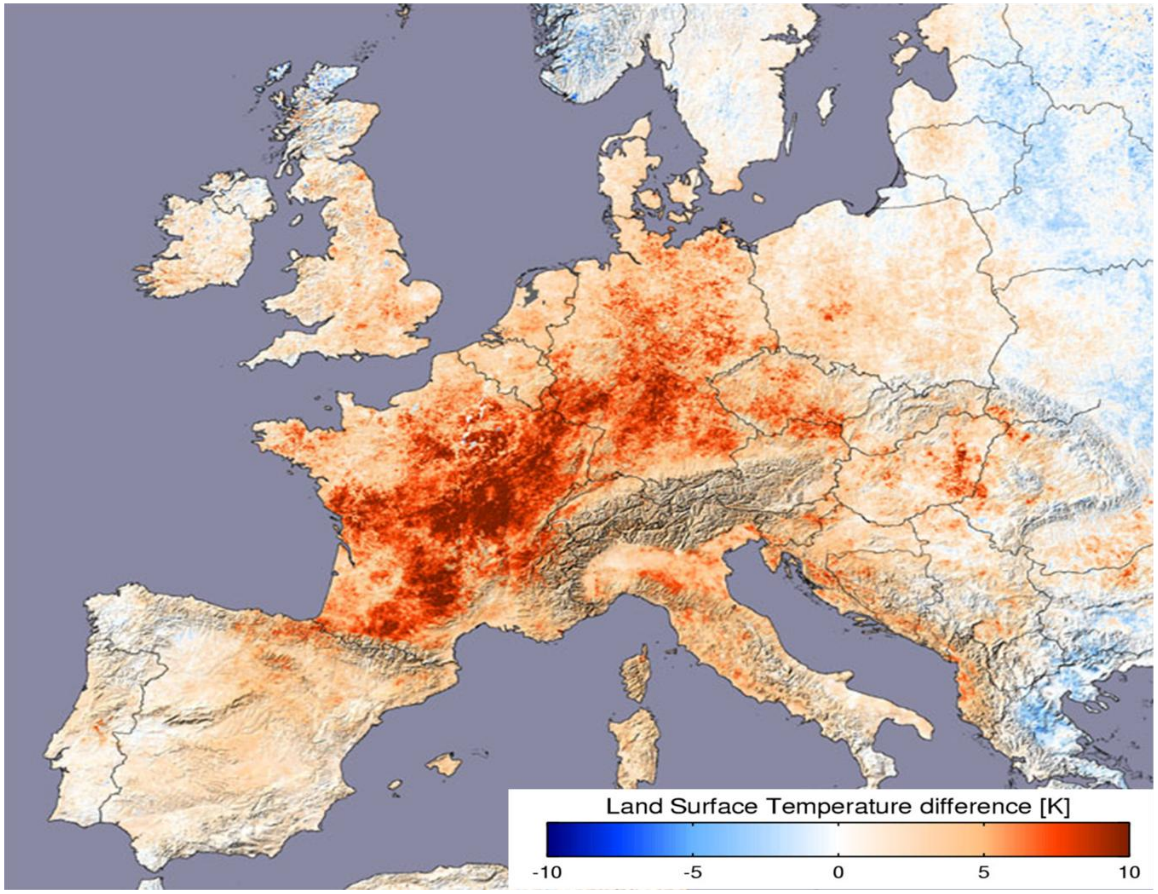
Here is an example to motivate the subject from the climate change perspective.

The year 2003 featured one of the hottest summers on record in southern Europe. Around 30,000 people are estimated to have died (in many cases, because they failed to receive adequate treatment for heatstroke).

It's not possible for a single event like this to say whether it was "due to climate change". But we can ask the following question:

What is the probability that such an event would occur under present-day climate conditions? And what would that probability have been if there had been no greenhouse-gas-induced global warming?

A paper by Stott et al. in *Nature* addressed that question directly.



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]

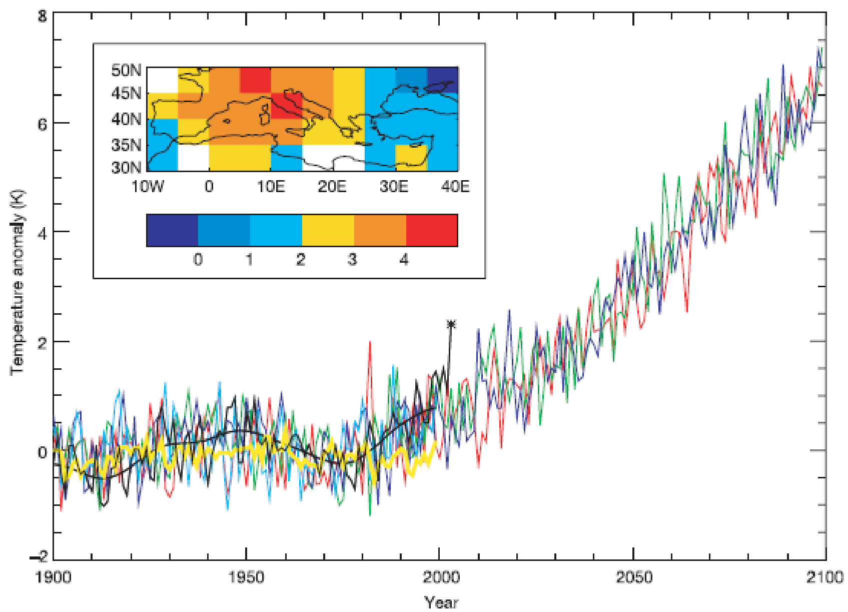


Figure 1 June–August temperature anomalies (relative to 1961–90 mean, in K) over the region shown in inset. Shown are observed temperatures (black line, with low-pass-filtered temperatures as heavy black line), modelled temperatures from four HadCM3 simulations including both anthropogenic and natural 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.

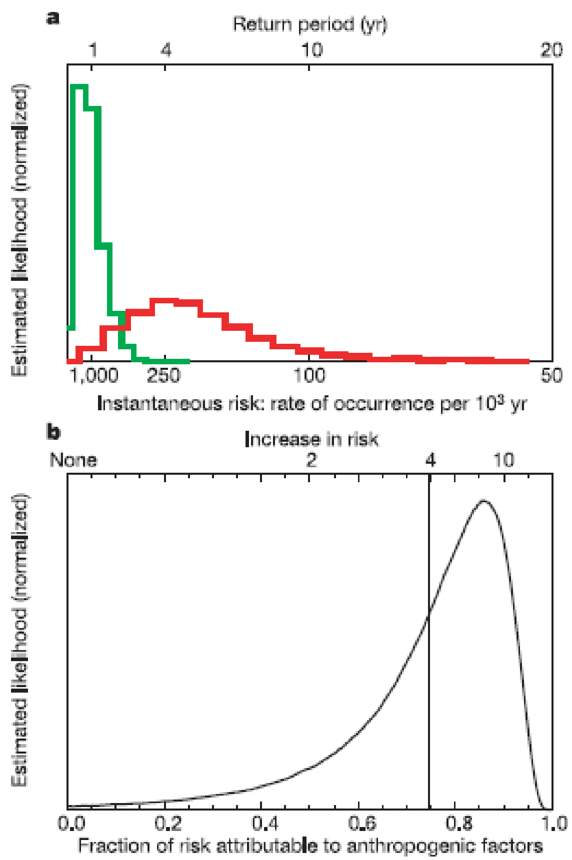


Figure 4 Change in risk of mean European summer temperatures exceeding the 1.6 K threshold. **a**, Histograms of instantaneous return periods under late-twentieth-century conditions in the absence of anthropogenic climate change (green line) and with anthropogenic climate change (red line). **b**, Fraction 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 PRESENTATION

I. Extreme value theory

- Probability Models
- Estimation
- Diagnostics

II. Example: North Atlantic Storms

III. Example: European Heatwave

IV. Example: Wang Junxia's Record

V. Example: Insurance Extremes

VI. Example: Trends in Extreme Rainfall Events

VII. Multivariate Extremes and Max-stable Processes

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

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

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)

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

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.

POINT PROCESS APPROACH

Homogeneous case:

Exceedance $y > u$ at time t has probability

$$\frac{1}{\psi}\left(1+\xi\frac{y-\mu}{\psi}\right)_+^{-1/\xi-1}\exp\left\{-\left(1+\xi\frac{u-\mu}{\psi}\right)_+^{-1/\xi}\right\}dydt$$

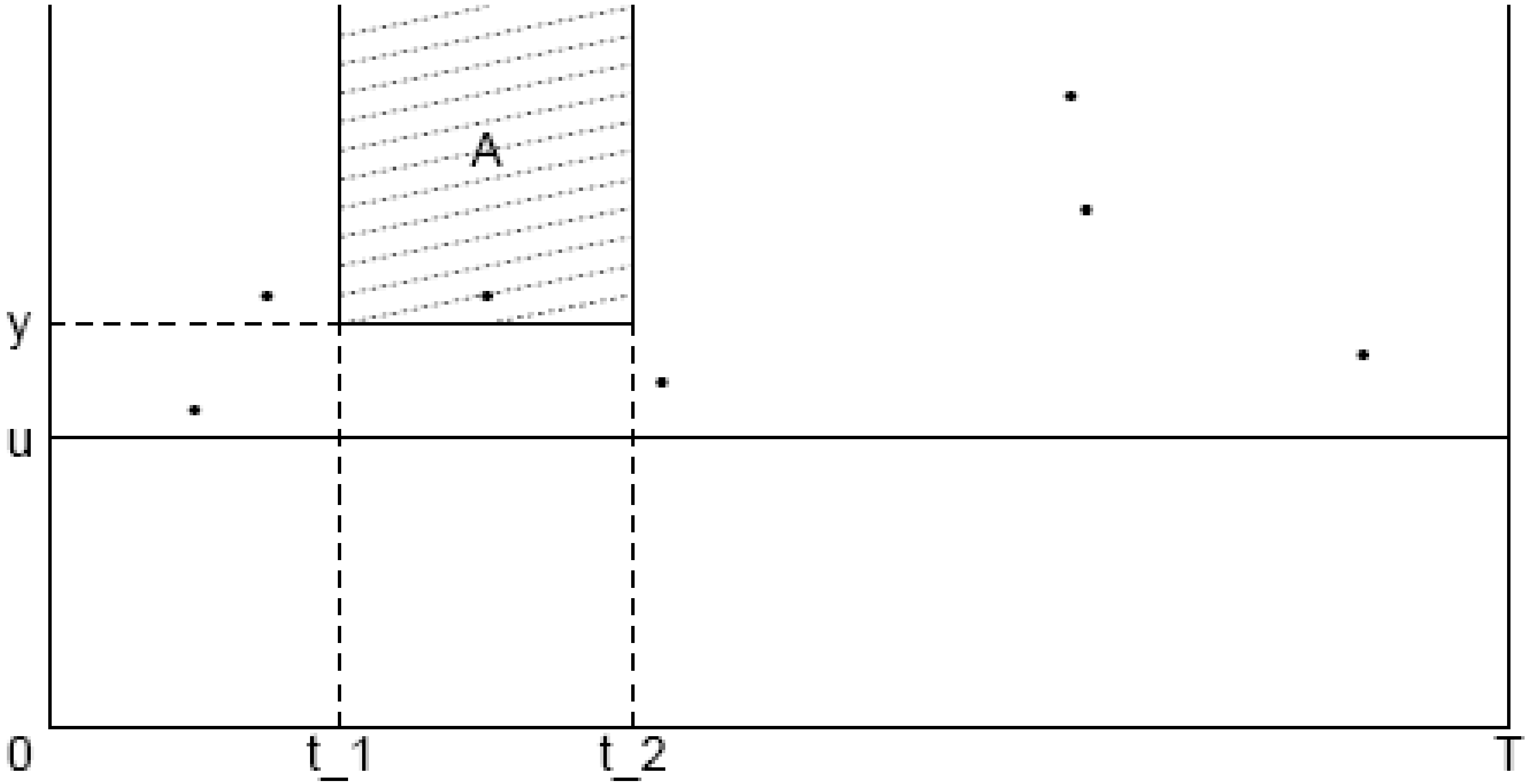


Illustration of point process model.

Inhomogeneous case:

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

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

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

- Estimation by maximum likelihood

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.

DIAGNOSTICS

Gumbel plots

QQ plots of residuals

Mean excess plot

Z and W plots

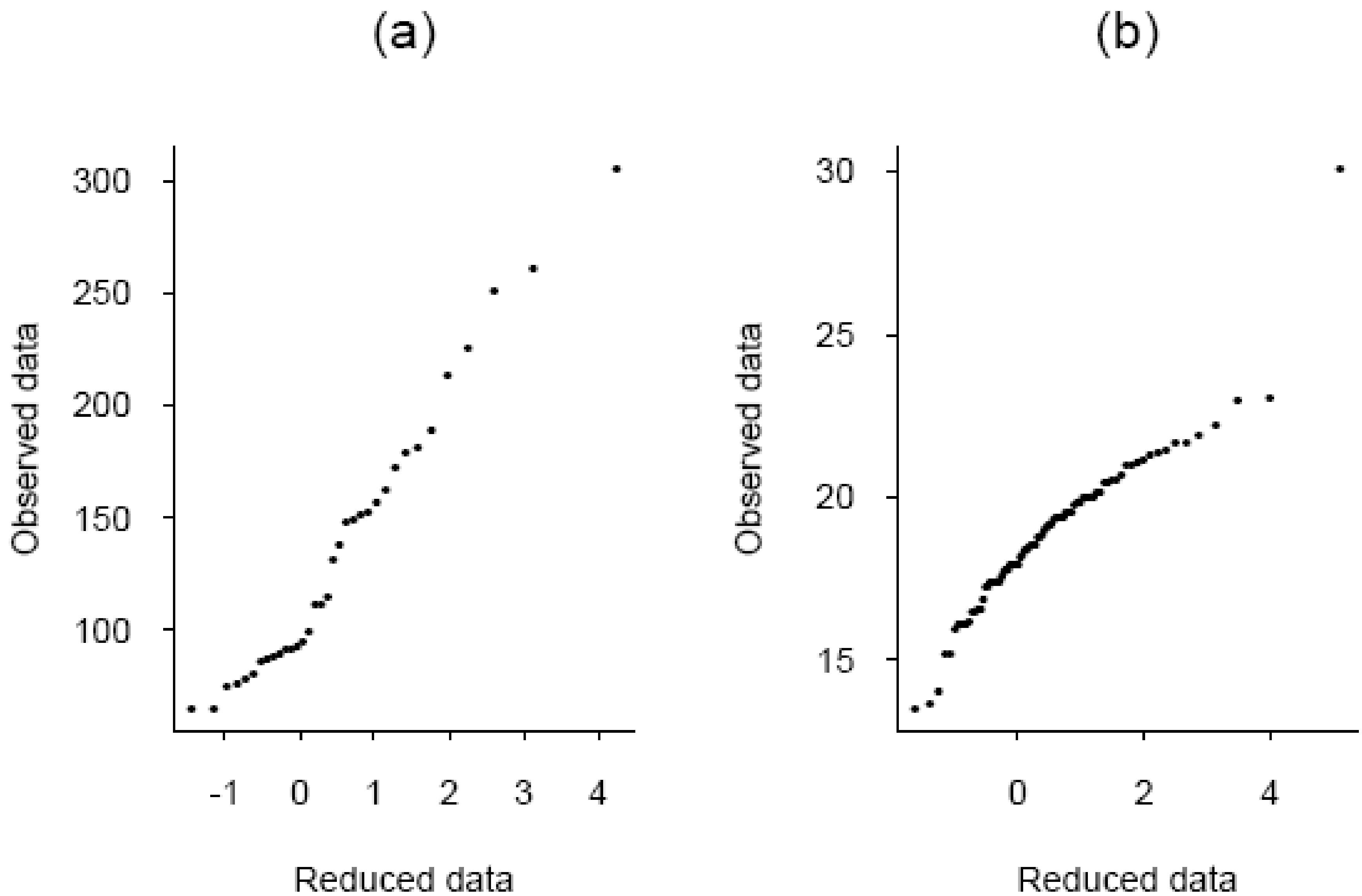
Gumbel plots

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.

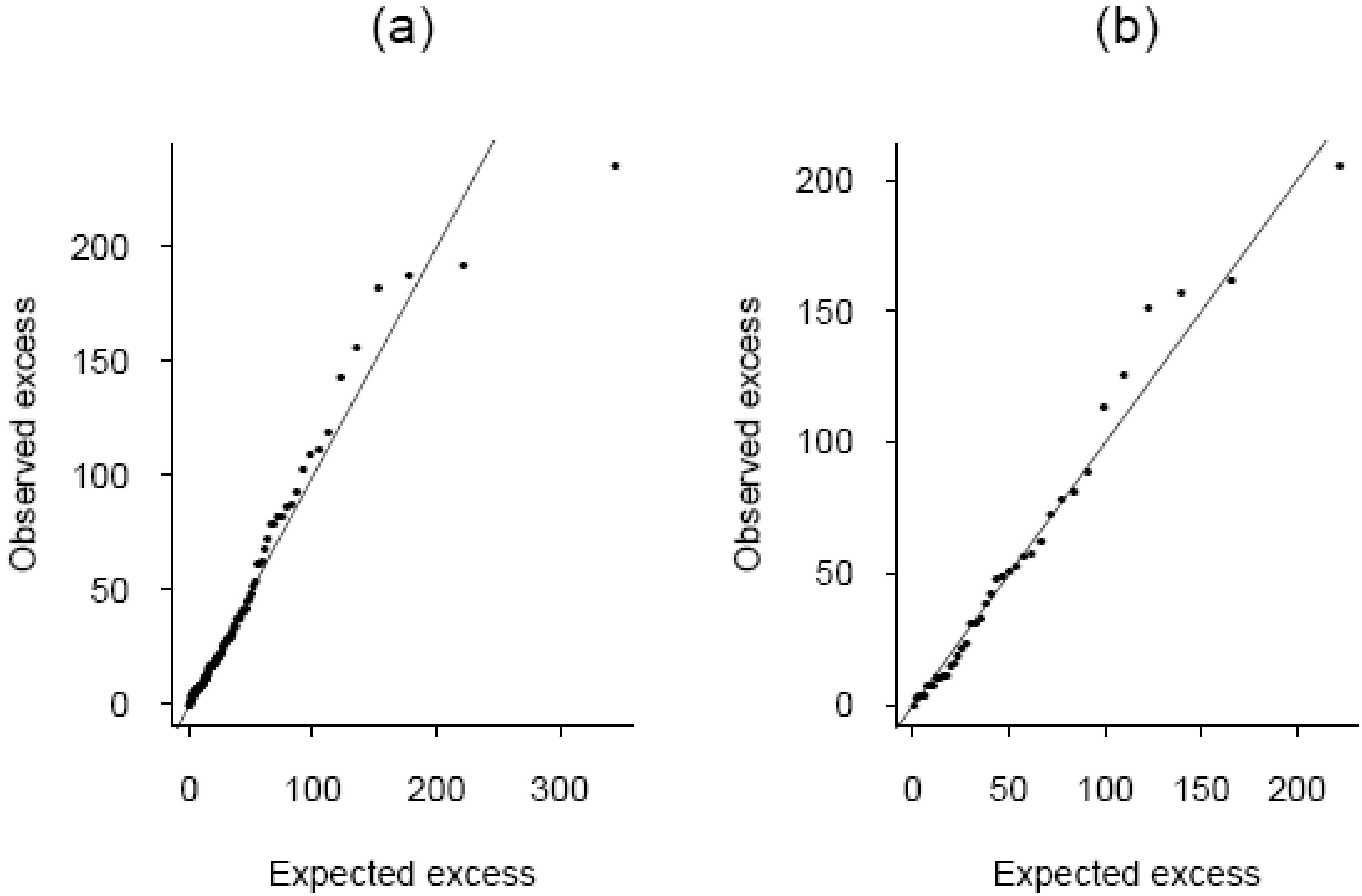
QQ plots of residuals

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...



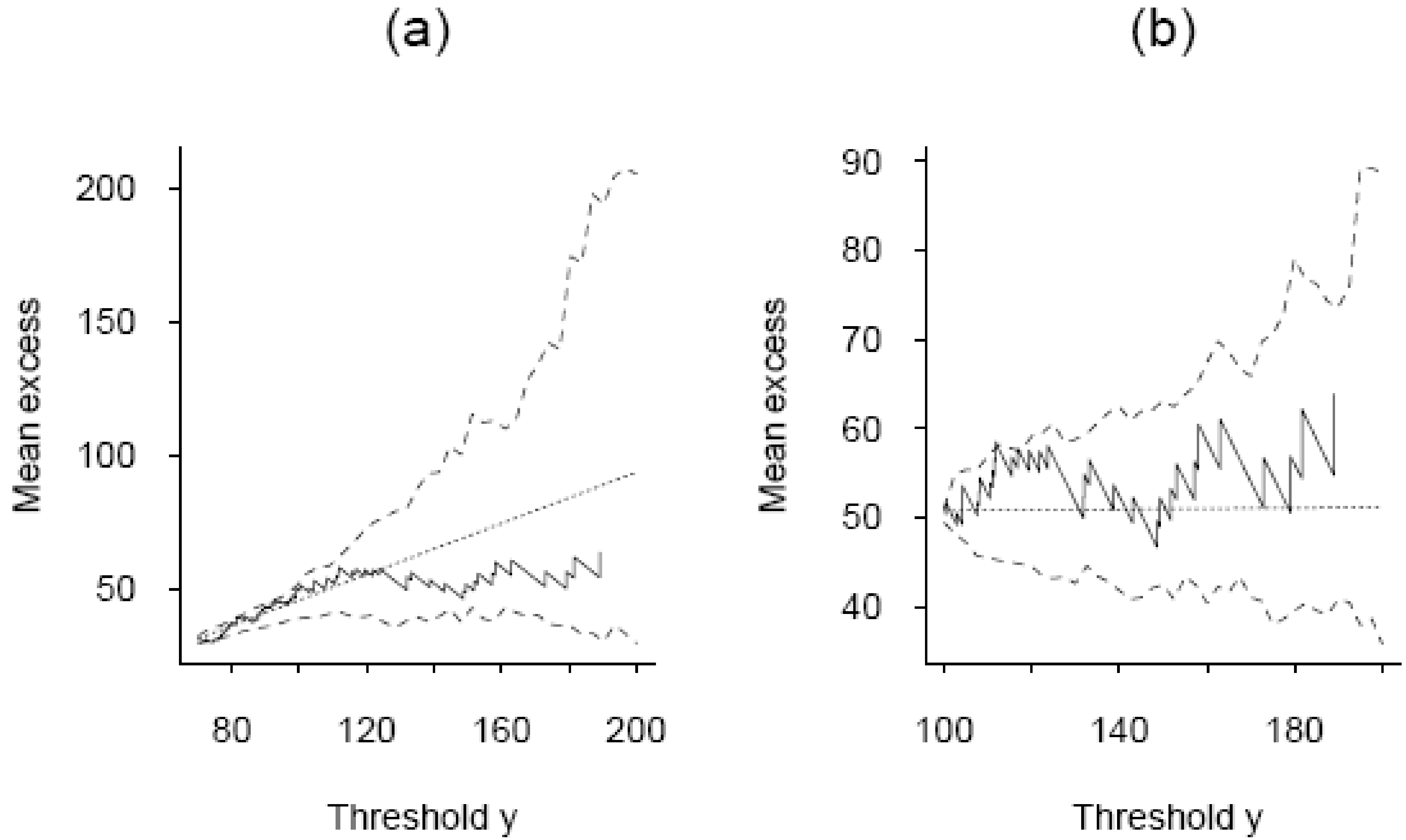
QQ plots for GPD, Nidd data. (a) $u = 70$. (b) $u = 100$.

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).

Z- and W-statistic plots

Consider nonstationary model with μ_t, ψ_t, ξ_t dependent on t.

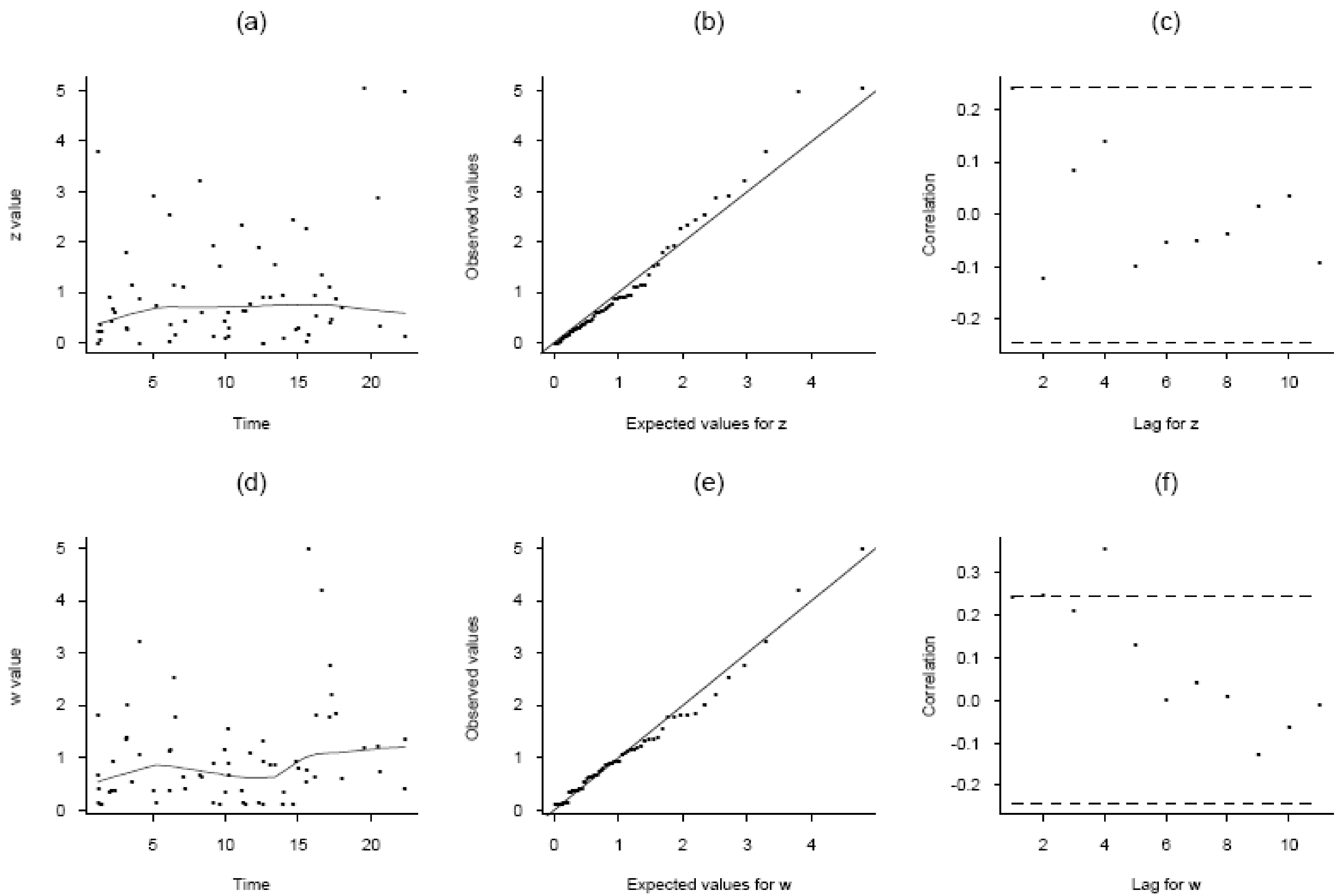
Z statistic based on intervals between exceedances T_k:

$$
\begin{aligned}
Z_k &= \int_{T_{k-1}}^{T_k} \lambda_u(s)ds, \\
\lambda_u(s) &= \{1+\xi_s(u-\mu_s)/\psi_s)\}^{-1/\xi_s}.
\end{aligned}
$$

W statistic based on excess values: if Y_k is excess over threshold at time T_k,

$$
W_k = \frac{1}{\xi_{T_k}} \log\left\{1+\frac{\xi_{T_k}Y_k}{\psi_{T_k}+\xi_{T_k}(u-\mu_{T_k})}\right\}.
$$

Idea: if the model is exact, both Z_k and W_k and i.i.d. exponential with mean 1. Can test this with various plots.



Diagnostic plots based on Z and W statistics for Charlotte wind-speed data.

II. NORTH ATLANTIC CYCLONES

Data from HURDAT

Maximum windspeeds in all North Atlantic Cyclones from 1851–2007

TROPICAL CYCLONES FOR THE NORTH ATLANTIC

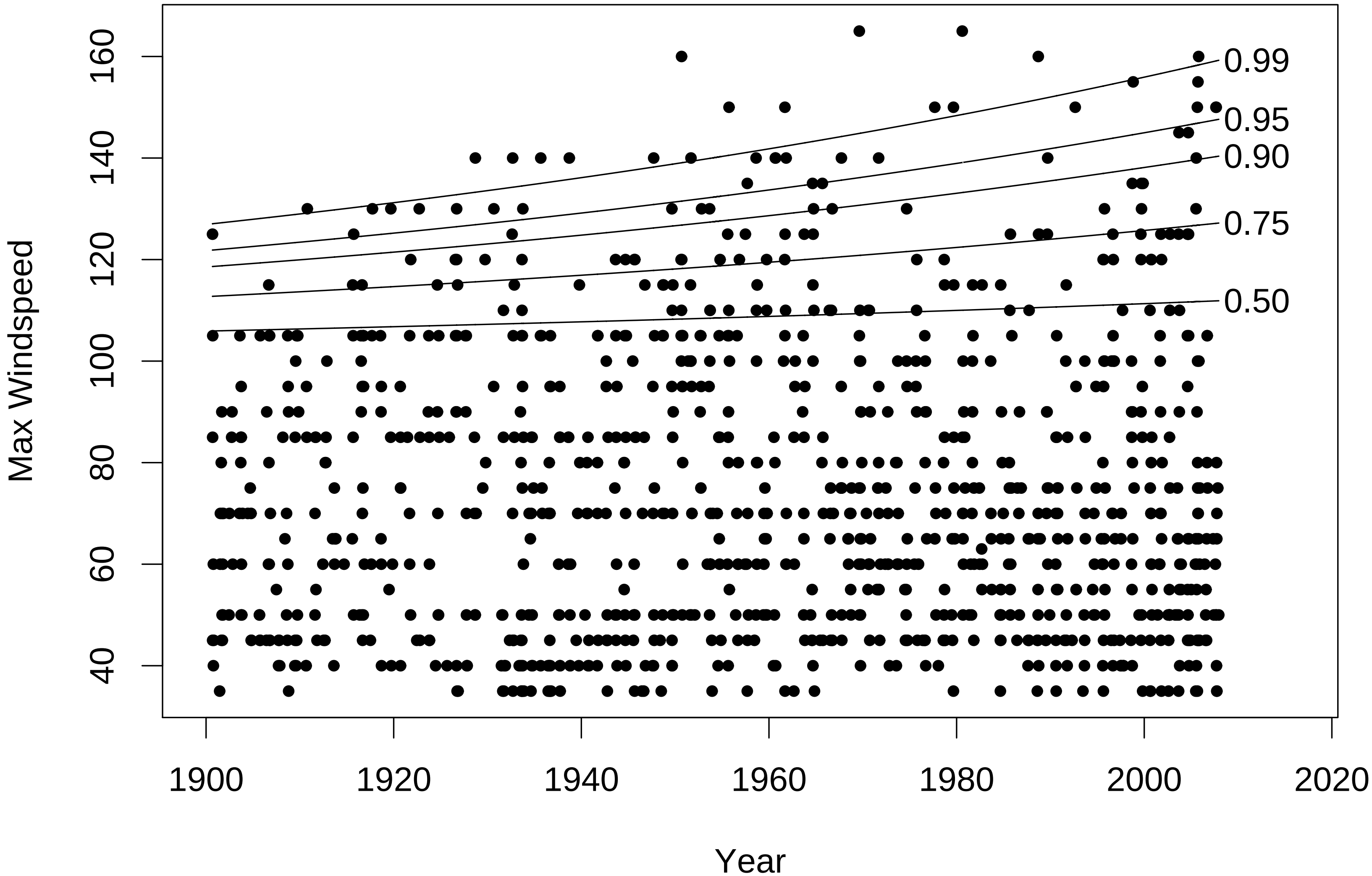
POT MODELS 1900–2007, u=102.5

Model	p	NLLH	NLLH+p
Gumbel	2	847.8	849.8
GEV	3	843.8	846.8
GEV, lin μ	4	834.7	838.7
GEV, quad μ	5	833.4	838.4
GEV, cubic μ	6	829.8	835.8
GEV, lin μ, lin $\log \psi$	5	828.0	833.0
GEV, quad μ, lin $\log \psi$	6	826.8	832.8
GEV, lin μ, quad $\log \psi$	6	827.2	833.2

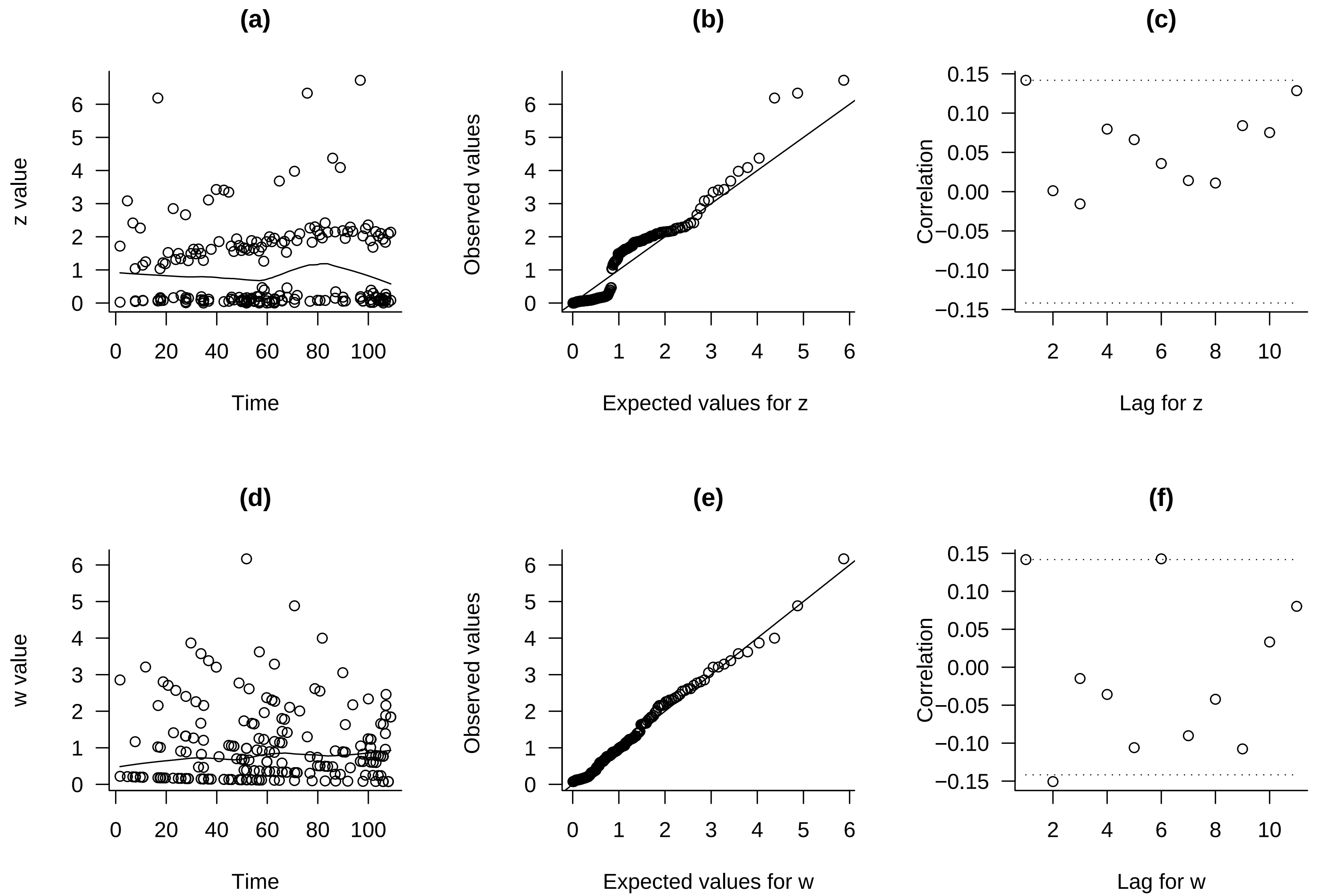
Fitted model: $\mu = \beta_0 + \beta_1 t,\ \log \psi = \beta_2 + \beta_3 t,\ \xi$ const

	β_0	β_1	β_2	β_3	ξ
Estimate	102.5	0.0158	2.284	0.0075	−0.302
S.E.	2.4	0.049	0.476	0.0021	0.066

TROPICAL CYCLONES FOR THE NORTH ATLANTIC
Fitted Max Windspeed Quantiles for 1900−2007



Diagnostic Plots for Atlantic Cyclones



III. EUROPEAN HEATWAVE

Data:

5 model runs from CCSM 1871–2100, including anthropogenic forcing

2 model runs from UKMO 1861–2000, including anthropogenic forcing

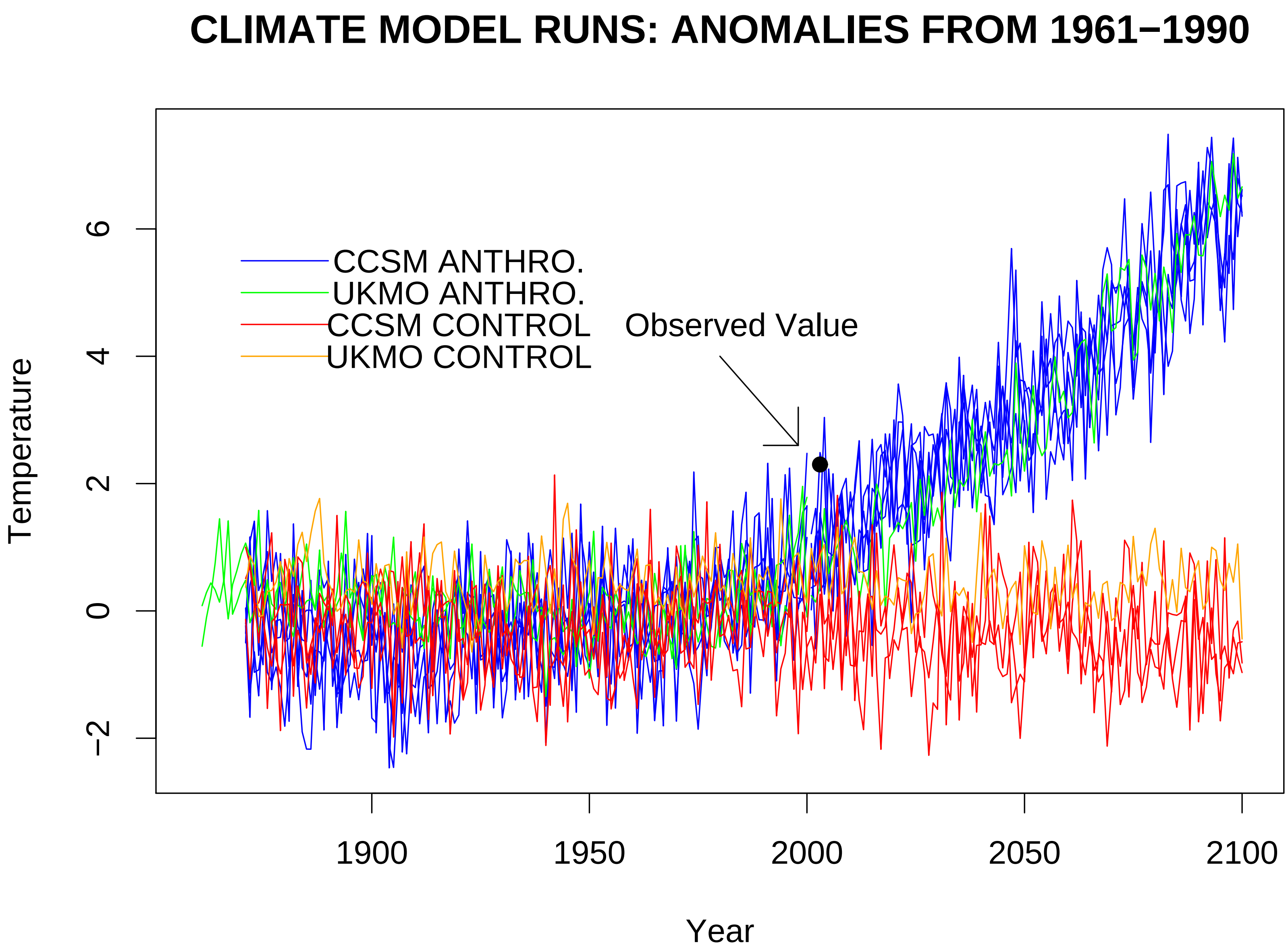
1 model runs from UKMO 2001–2100, including anthropogenic forcing

2 control runs from CCSM, 230+500 years

2 control runs from UKMO, 341+81 years

All model data have been calculated for the grid box from 30–50° N, 10° W–40° E, annual average temperatures over June–August

Expressed an anomalies from 1961–1990, similar to Stott, Stone and Allen (2004)


CLIMATE MODEL RUNS: ANOMALIES FROM 1961−1990
CCSM ANTHRO.
UKMO ANTHRO.
CCSM CONTROL
UKMO CONTROL
Observed Value
Temperature
−2
0
2
4
6
1900
1950
2000
2050
2100
Year

Method:

Fit POT models with various trend terms to the anthropogenic model runs, 1861–2010

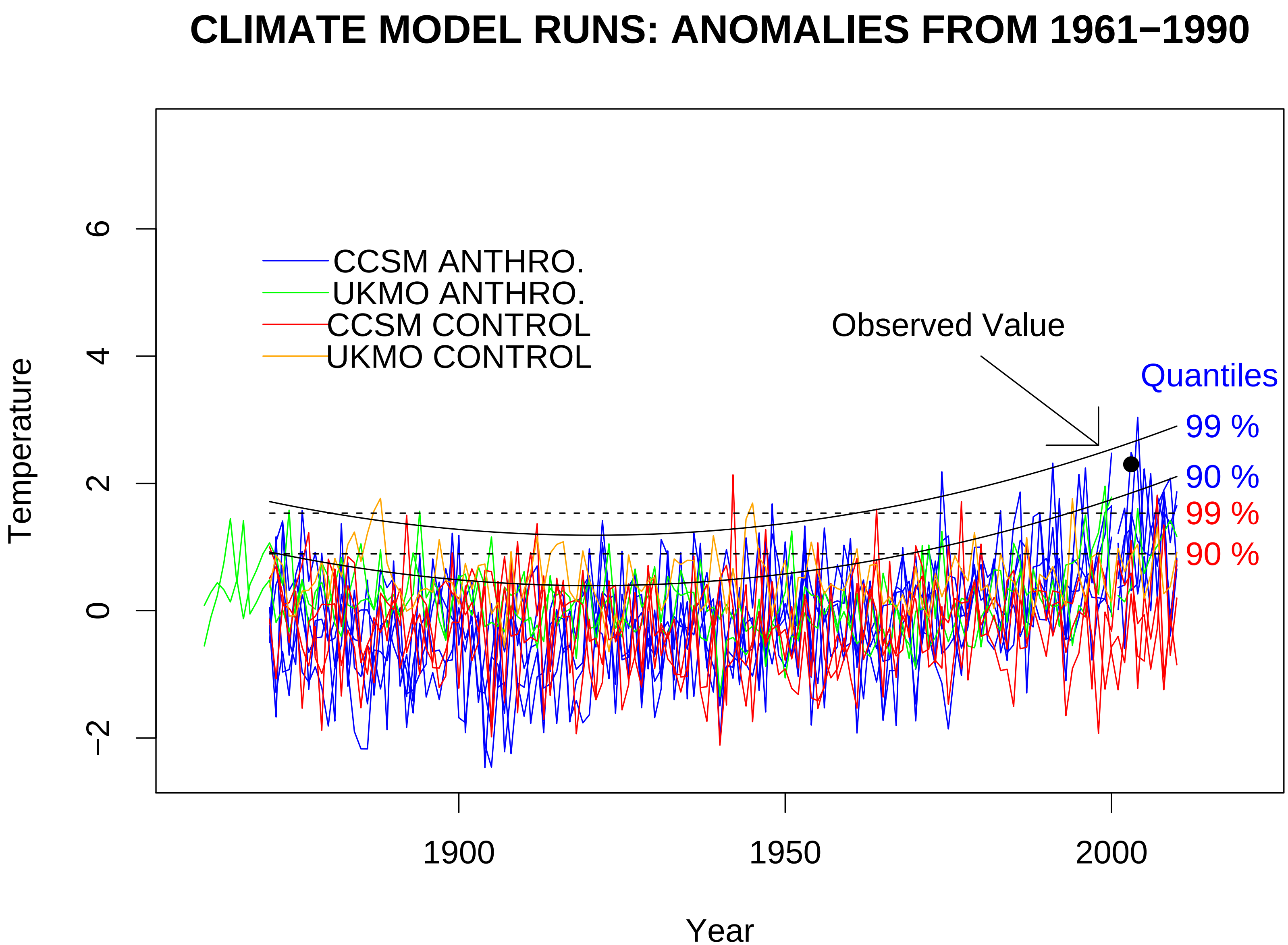
Also fit trend-free model to control runs ($\mu = 0.176$, $\log\psi = -1.068$, $\xi = -0.068$)

POT MODELS 1861–2010, u=1

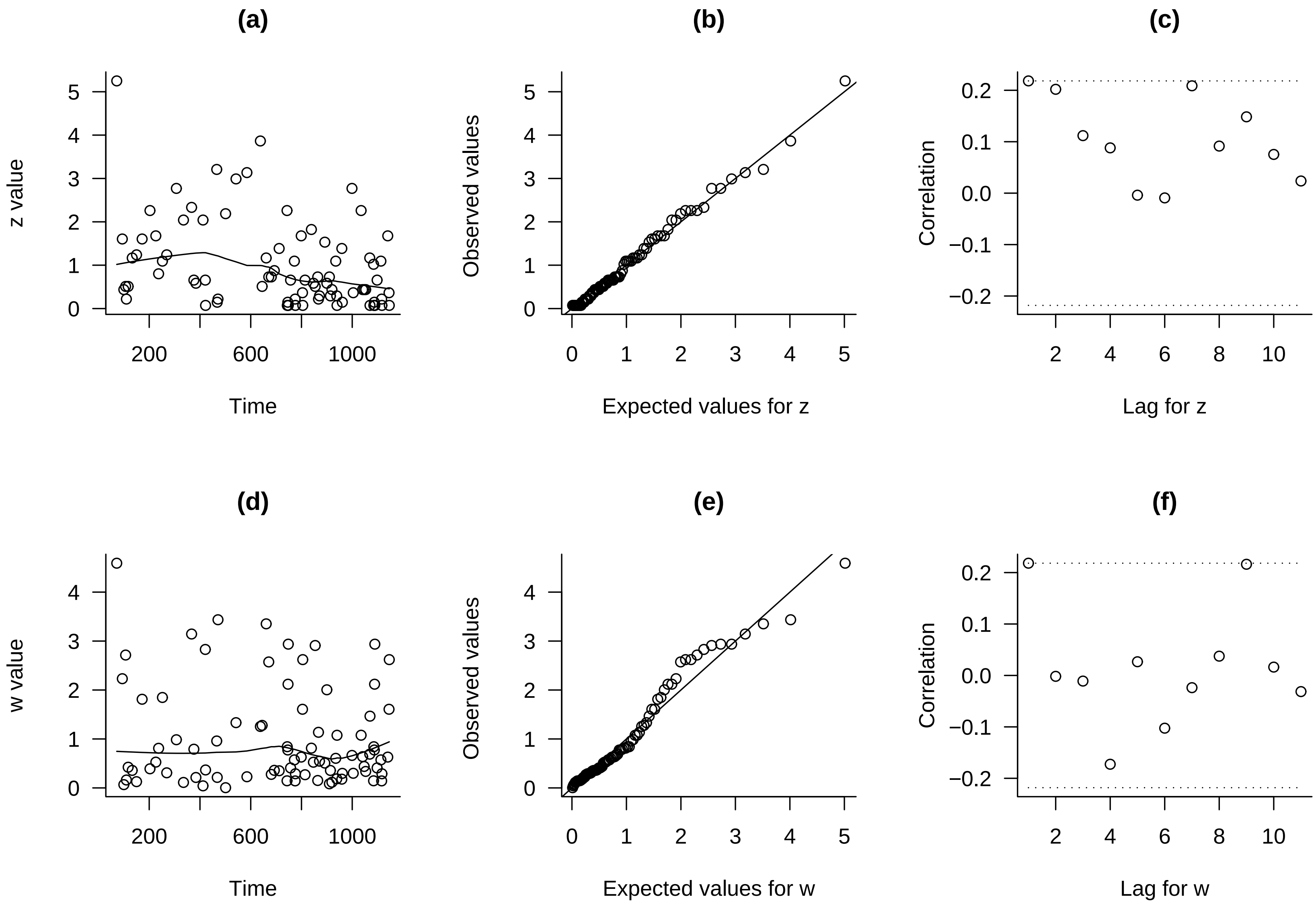
Model	p	NLLH	NLLH+p
Gumbel	2	349.6	351.6
GEV	3	348.6	351.6
GEV, lin μ	4	315.5	319.5
GEV, quad μ	5	288.1	293.1
GEV, cubic μ	6	287.7	293.7
GEV, quart μ	7	285.1	292.1
GEV, quad μ, lin $\log\psi$	6	287.9	293.9
GEV, quad μ, quad $\log\psi$	7	287.0	294.9

Fitted model: $\mu = \beta_0 + \beta_1 t + \beta_2 t^2, \ \psi, \xi$ const

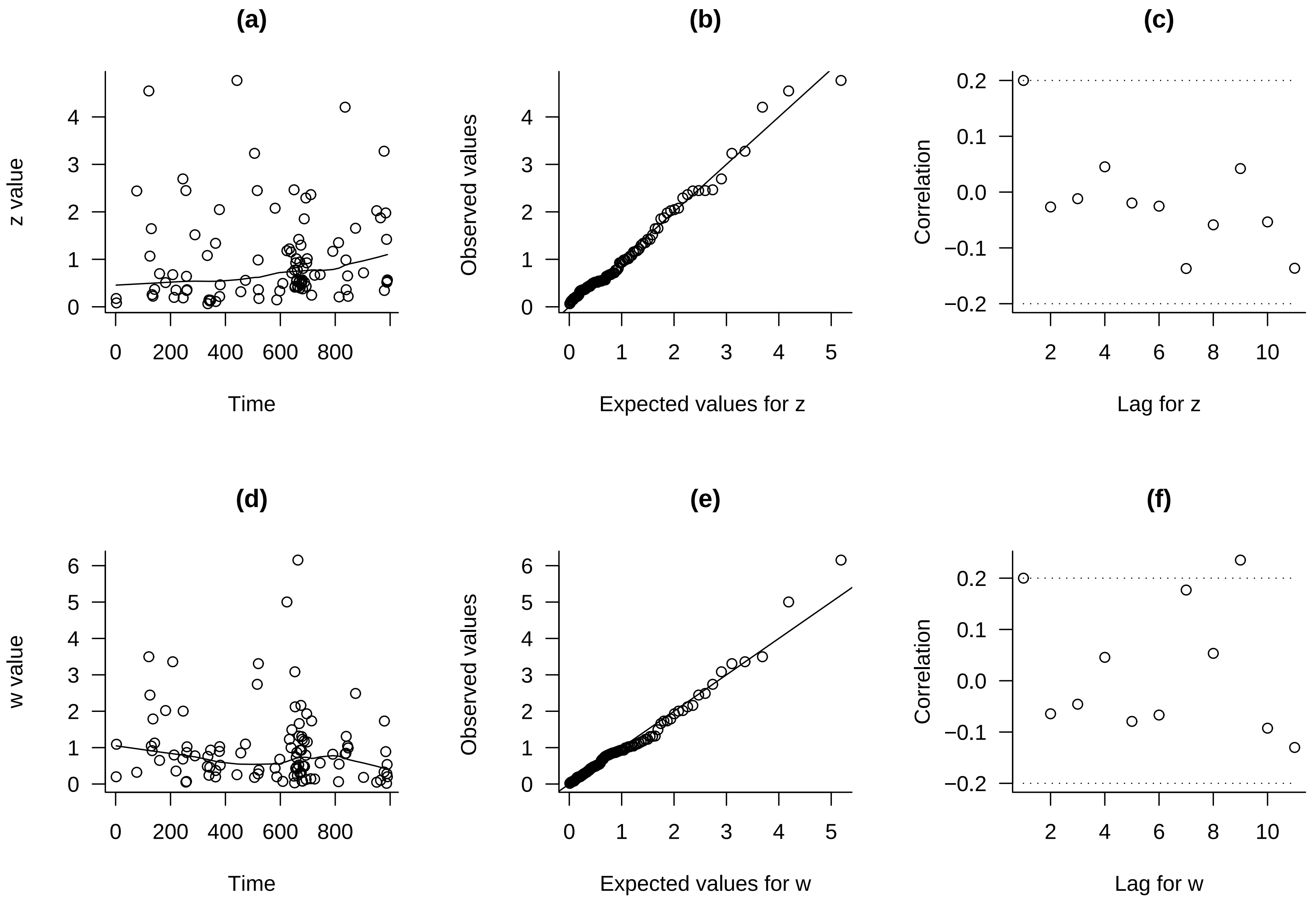
	β_0	β_1	β_2	$\log\psi$	ξ
Estimate	–0.187	–0.030	0.000215	0.047	0.212
S.E.	0.335	0.0054	0.00003	0.212	0.067


CLIMATE MODEL RUNS: ANOMALIES FROM 1961−1990
CCSM ANTHRO.
UKMO ANTHRO.
CCSM CONTROL
UKMO CONTROL
Observed Value
Quantiles
99 %
90 %
99 %
90 %
Temperature
−2
0
2
4
6
1900
1950
2000
Year

Diagnostic Plots for Temperatures (Control)



Diagnostic Plots for Temperatures (Anthropogenic)



We now estimate the probabilities of crossing various thresholds in 2003.

Express answer as N=1/(exceedance probability)

Threshold 2.3:

N=3024 (control), N=29.1 (anthropogenic)

Threshold 2.6:

N=14759 (control), N=83.2 (anthropogenic)

Comments

1. Differences from Stott-Stone-Allen conclusions arise for two reasons; different settings of climate models (in particular, I used "control runs" where they used "natural forcings"); but also, different climate models. Not clear how much of the difference is due to inconsistencies between different climate models. (In particular, use of *anomalies* from 1961–1990 is very important, as the different models differ quite a bit in actual projected temperatures.)

2. A more sophisticated approach would be to use observational data from 1850–2002 to calibrate the climate models; rather similar to the approach of Tebaldi *et al.* (recall opening classes of this course). This is an open research problem!

IV. WANG JUNXIA'S RECORD

References:

Robinson and Tawn (1995), *Applied Statistics*

Smith (1997), *Applied Statistics*


ITALIA
Atlanta 1996
3438
Atlanta 1996
3154

On September 13, 1993, the Chinese athlete Wang Junxia ran a new world record for the 3000 meter track race in 8 minutes, 6.11 seconds (486.11 sec.).

This improved by 6.08 seconds the record she had run the day before, which in turn improved by 10.43 seconds on the previous existing record (502.62 seconds). Both times, she tested negatively for drugs.

Nevertheless there was widespread suspicion that the record was, indeed, drug assisted.

Robinson and Tawn compiled a record of the 5 independent best times each year from 1972–1992, together with Wang's record.

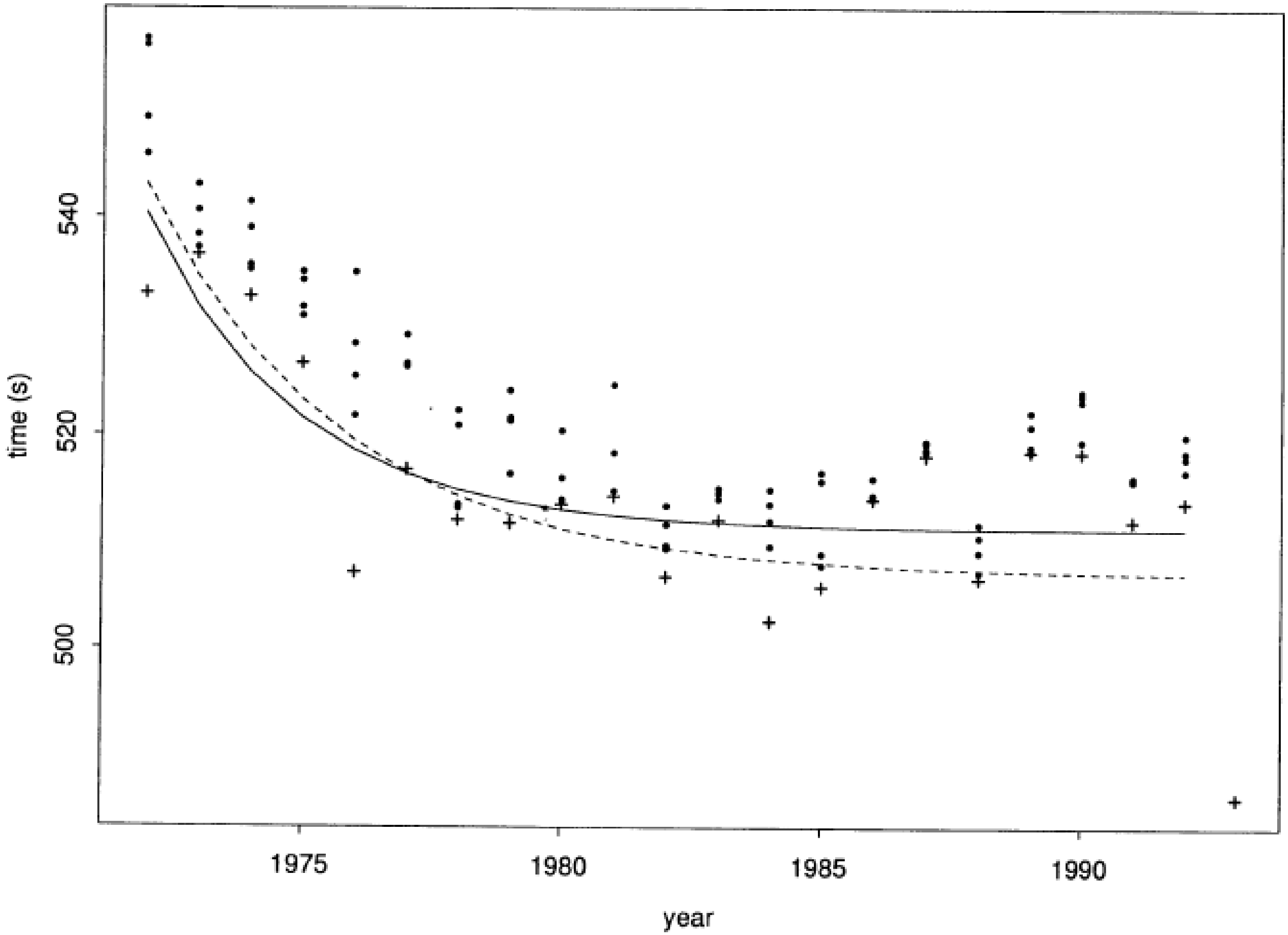


Fig. 1. Five best annual times for the 3000 m: +, annual minima; ———, regression function (1.5), estimated by using model A of Table 1; --------, regression function (1.5), estimated by using model C of Table 1

Initial models for the annual minimum:

$$
\begin{aligned}
G_t(x) &= 1 - \exp\left[-\left\{ 1 - \frac{\xi(x-\mu_t)}{\sigma} \right\}_+^{-1/\xi} \right], \\
\mu_t &= \alpha - \beta(1 - e^{-\gamma t}).
\end{aligned}
$$

Extended to likelihood based on 5 smallest values per year.

Based on this define

$$
x_{ult} = \begin{cases} \alpha - \beta + \frac{\sigma}{\xi} & \text{if } \xi < 0, \\ -\infty & \text{if } \xi \geq 0. \end{cases}
$$

Profile likelihood for x_{ult}: 90% confidence interval based on data up to 1992 is $(0, 501.1)$.

Limiting joint density of r (independent) minima per year:

Suppose the r smallest running times are $x_1 \le x_2 \le ... \le x_r$.

Then the joint density is

$$\frac{1}{\sigma^r}\left\{\prod_{i=1}^{r}\left(1-\xi\frac{x_i-\mu_t}{\sigma}\right)_+^{-1/\xi-1}\right\}\exp\left\{-\left(1-\xi\frac{x_r-\mu_t}{\sigma}\right)_+^{-1/\xi}\right\}$$

TABLE 1

Parameter estimates and standard errors for the GEV distribution fitted to four models†

Parameter	*Estimates for the following models:*			
	A	*B*	*C*	*D*
α	555 (11)	558 (4)	556 (4)	559 (4)
β	41.2 (10.2)	48.7 (3.8)	47.2 (3.5)	48.3 (3.5)
γ	0.343 (0.089)	0.272 (0.034)	0.264 (0.029)	0.296 (0.033)
σ	4.73 (1.00)	4.85 (0.40)	4.75 (0.25)	4.41 (0.25)
ξ	−0.103 (0.269)	−0.197 (0.068)	−0.254 (0.045)	−0.298 (0.056)
x_{ult}	467.41 (115.22)	484.37 (9.78)	489.91 (4.14)	492.66 (3.49)

† Model A—women's 3000 m annual minima data (1972–92); model B—women's 3000 m, five independent best annual times (1972–92); model C—women's 1500 m and 3000 m, five independent best annual times (1972–92); model D—women's 1500 m and 3000 m, five independent best annual times (1972–92) incorporating the championship effect.

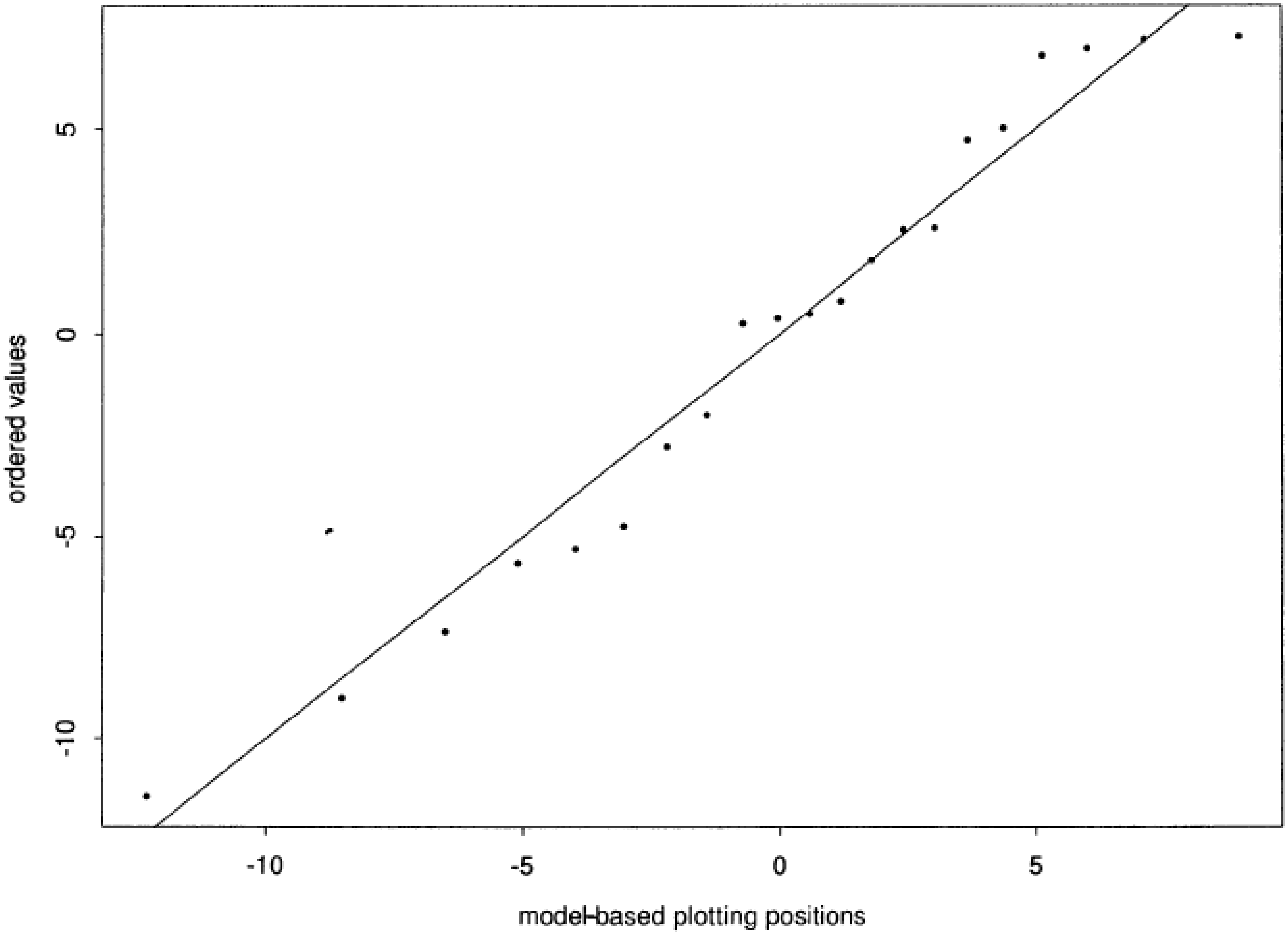


Fig. 2. Quantile–quantile plot of 3000 m residual annual minimum times: residuals that are negative correspond to the lower tail of interest

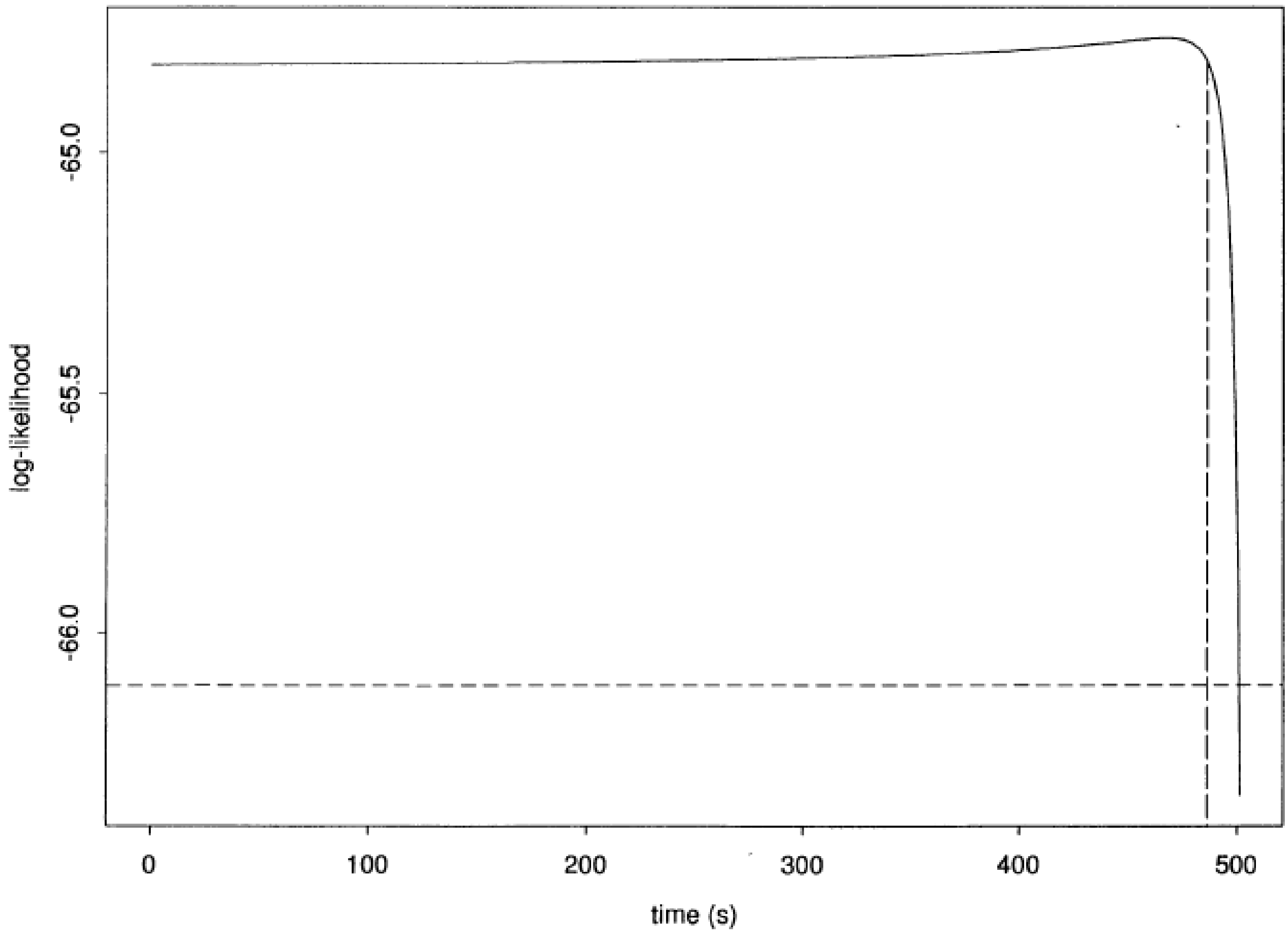


Fig. 3. Profile log-likelihood for x_{ult} as estimated using 3000 m annual minima: ----, 90% confidence interval; ¦, Wang's time

Smith (1997) suggested instead fitting data form 1980–1992 without trend.

In this model, a 90% confidence interval for x_{ult} was (488.2,502.29) (excluding Wang's record)

A 95% confidence interval for x_{ult} was (481.9,502.43) (including Wang's record)

One-sided P-value .039

Still not conclusive

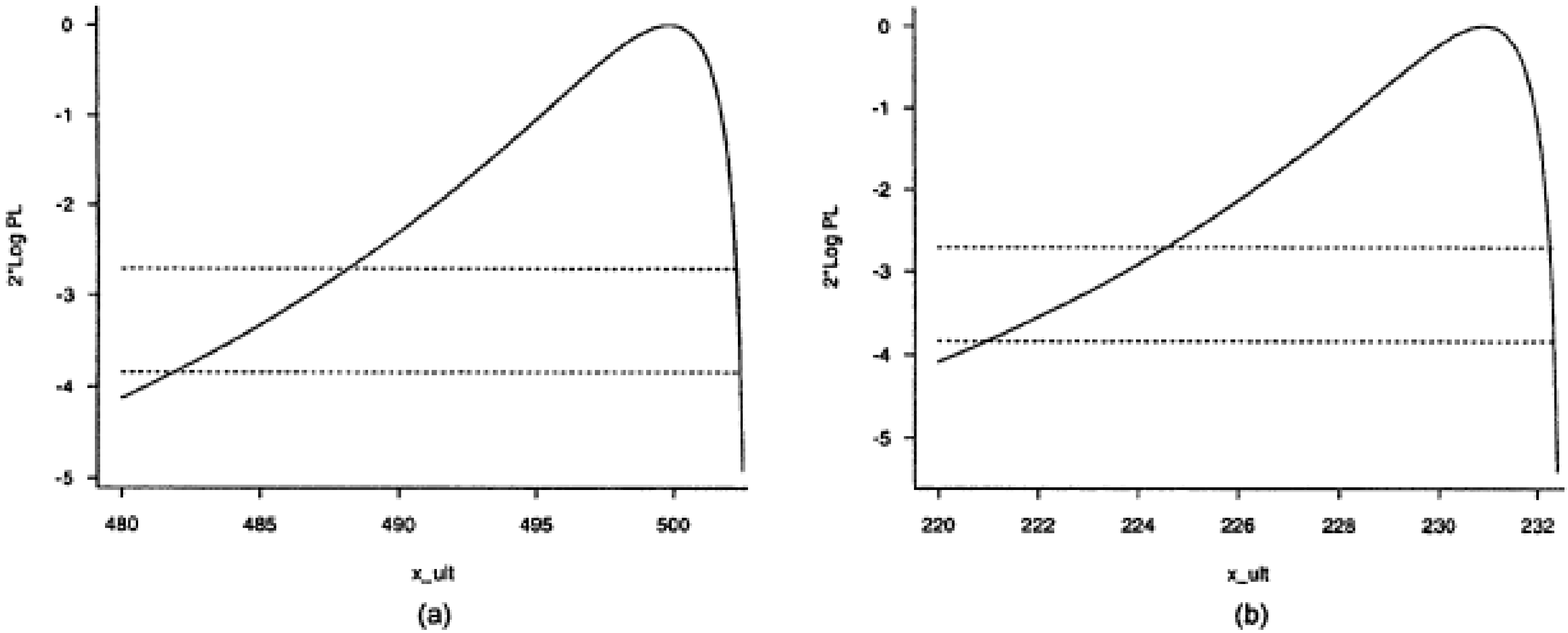


Fig. 1. Profile likelihood plots: (a) 3000 m data; (b) 1500 m data

Alternative Bayesian Analysis

Compute the *predictive distribution* for the record in 1993, *conditional on a new record being set*

Use prior density $\pi(\mu, \sigma, \xi) \propto \sigma^{-1}$ on $-\infty < \mu < \infty$, $\sigma > 0$, $\xi < 0$.

If $\mathbf{Y}_t$ denote observed (prior) data up to year t, y_{min} is existing record, and the annual minimum for year $t+1$ has c.d.f. $G_{t+1}(y; \mu, \sigma, \xi)$, then the conditional predictive c.d.f. for the record y in year $t+1$, given $y < y_{\text{min}}$, is

$$\int\int\int \frac{G_{t+1}(y; \mu, \sigma, \xi)}{G_{t+1}(y_{\text{min}}; \mu, \sigma, \xi)} \pi(\mu, \sigma, \xi \mid \mathbf{Y}_t) d\mu d\sigma d\xi.$$

Main Result

Posterior probability of observed record ≤ 486.11 is .00047 (subsequently revised to .0006; see Smith (2003)).

This seems to give much more convincing evidence that Wang's record was indeed a very unusual outlier, whether drug-assisted or not.

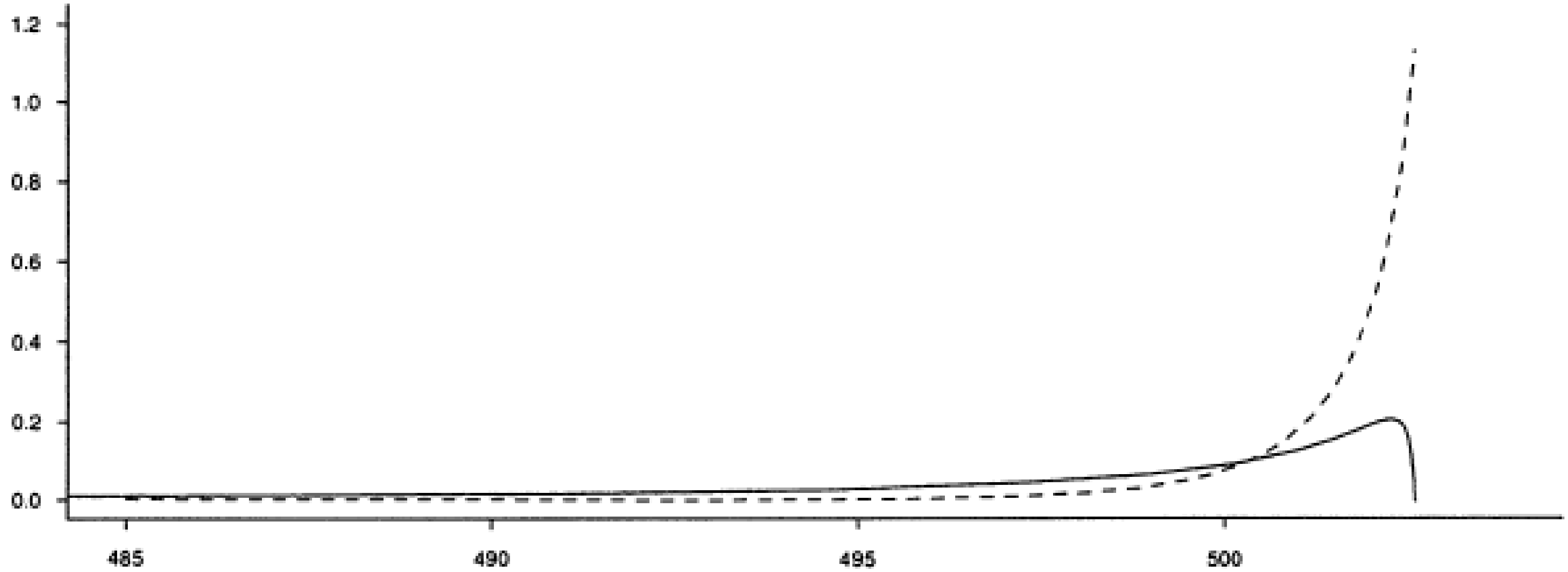


Fig. 2. Bayesian densities: ——— , posterior density; - - - - , predictive density

Posterior density for x_{ult} as well as conditional predictive density for new record. The latter has a much more rapidly decreasing left tail.

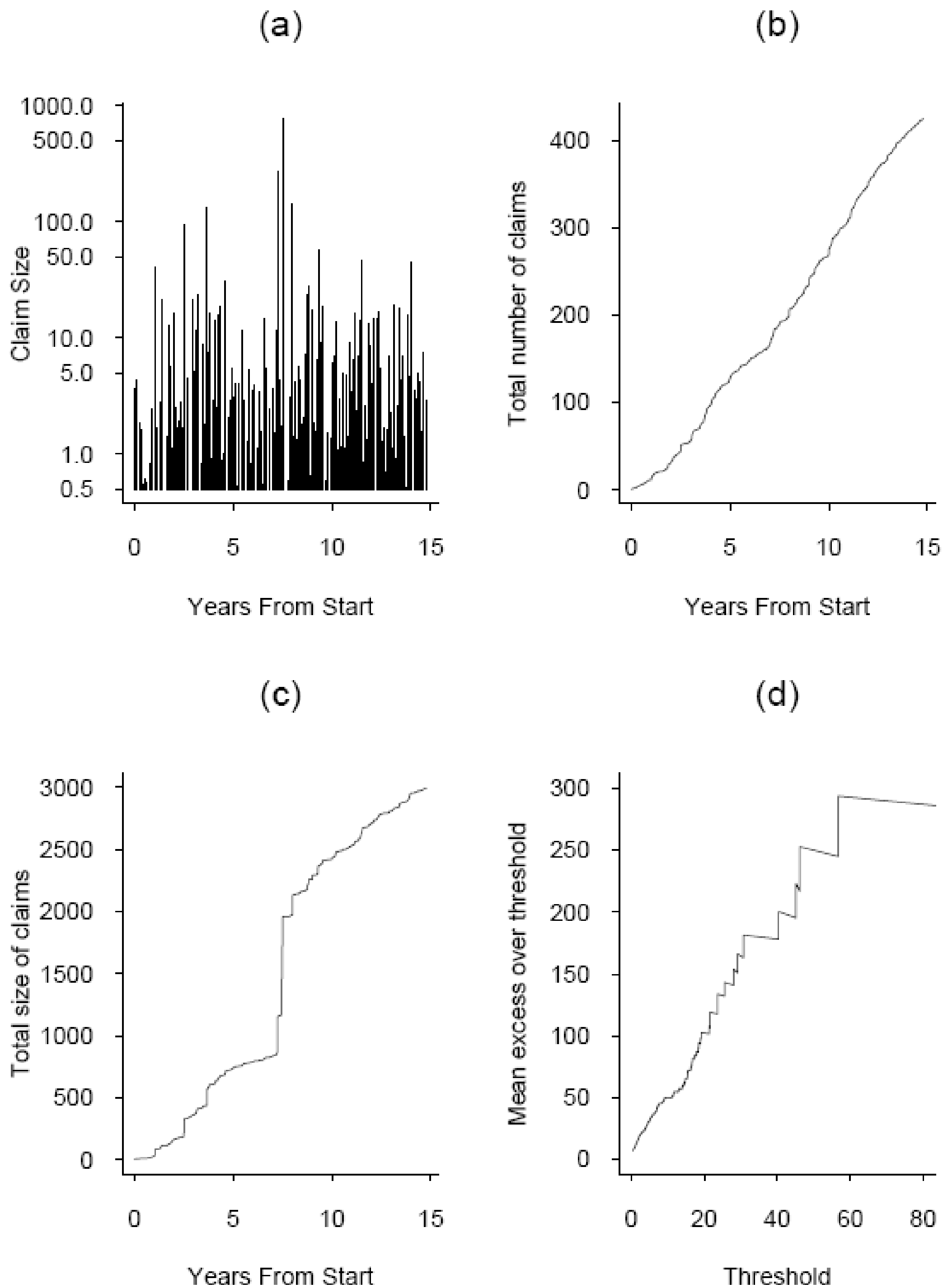
Take-Home Message

For computing the probability of a specific event, a *predictive distribution* may be much more meaningful than a posterior or likelihood-based interval for some parameter. This leads naturally into Bayesian approaches, because only a Bayesian approach adequately takes into account *both* the uncertainty of the unknown parameters *and* the randomness of the outcome itself.

Next, we see an example of the same reasoning, applied in the context of an insurance risk problem.

V. INSURANCE EXTREMES

(joint work with Dougal Goodman)



Oil company data:

(a) Main dataset

(b) cumulative number of claims

(c) cumulative total claims

(d) mean excess plot

Main dataset: 393 insurance claims experienced by a large oil company, subdivided into 6 classes of claim

Objective to characterize risk to the company from future extreme claims

Initial analysis: fit GPD and point process model to exceedances over various thresholds.

u	N_u	Mean Excess	$\hat{\sigma}$	(S.E.)	$\hat{\xi}$	(S.E.)
0.5	393	7.11	1.02	(0.10)	1.01	(0.10)
2.5	132	17.89	3.47	(0.59)	0.91	(0.17)
5	73	28.9	6.26	(1.44)	0.89	(0.22)
10	42	44.05	10.51	(2.76)	0.84	(0.25)
15	31	53.60	5.68	(2.32)	1.44	(0.45)
20	17	91.21	19.92	(10.42)	1.10	(0.53)
25	13	113.7	33.76	(18.93)	0.93	(0.55)
50	6	37.97	150.8	(106.3)	0.29	(0.57)

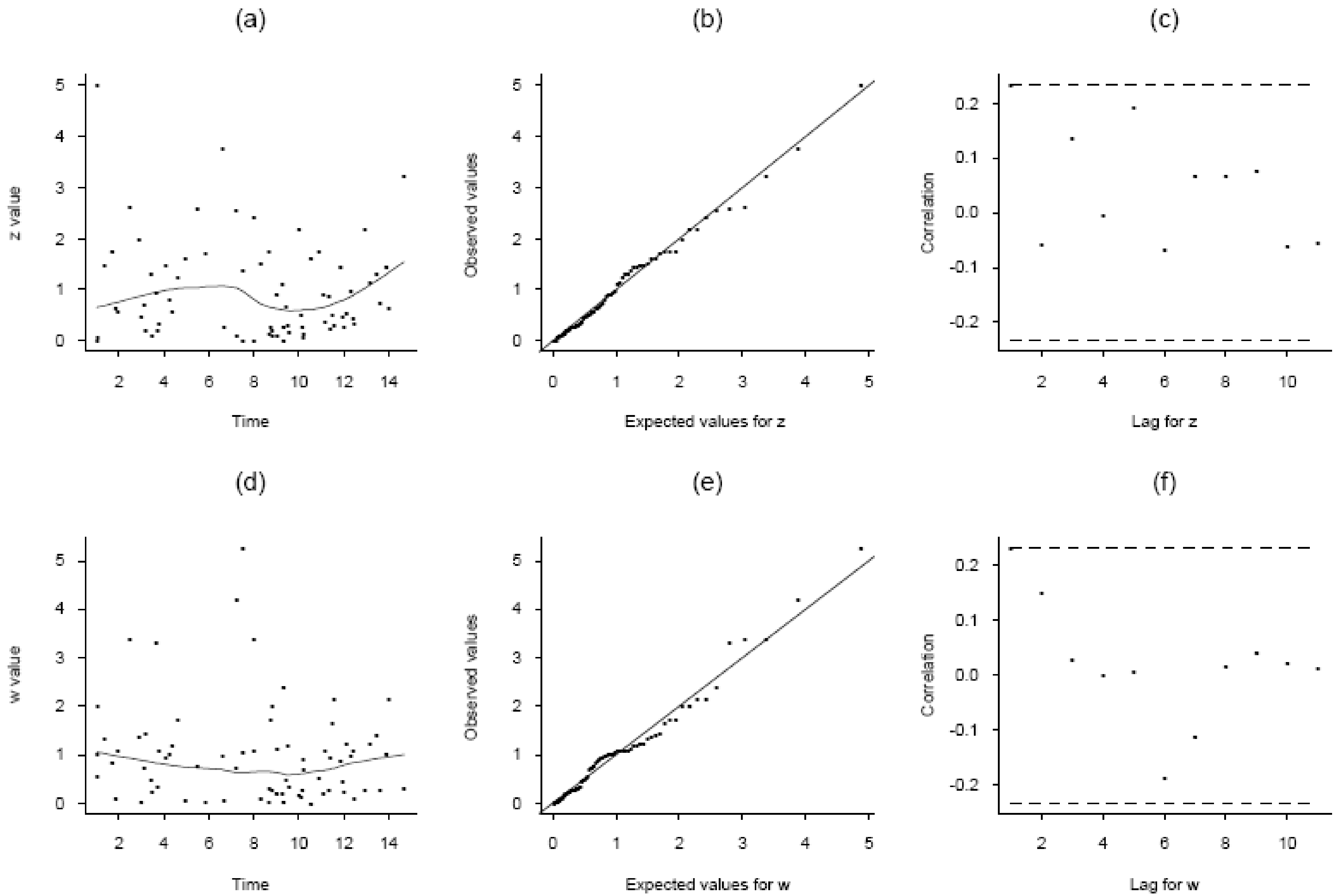
Table 10. Fit of the Generalised Pareto distribution to the excesses over various thresholds. The threshold is denoted u, and N_u the number of exceedances over u; we also tabulate the mean of all excesses over u and the maximum likelihood estimates of the GPD parameters σ and ξ. Standard errors are in parentheses.

Fitted parameters for GPD model

u	N_u	$\hat{\mu}$	(S.E.)	$\log \hat{\psi}$	(S.E.)	$\hat{\xi}$	(S.E.)
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)

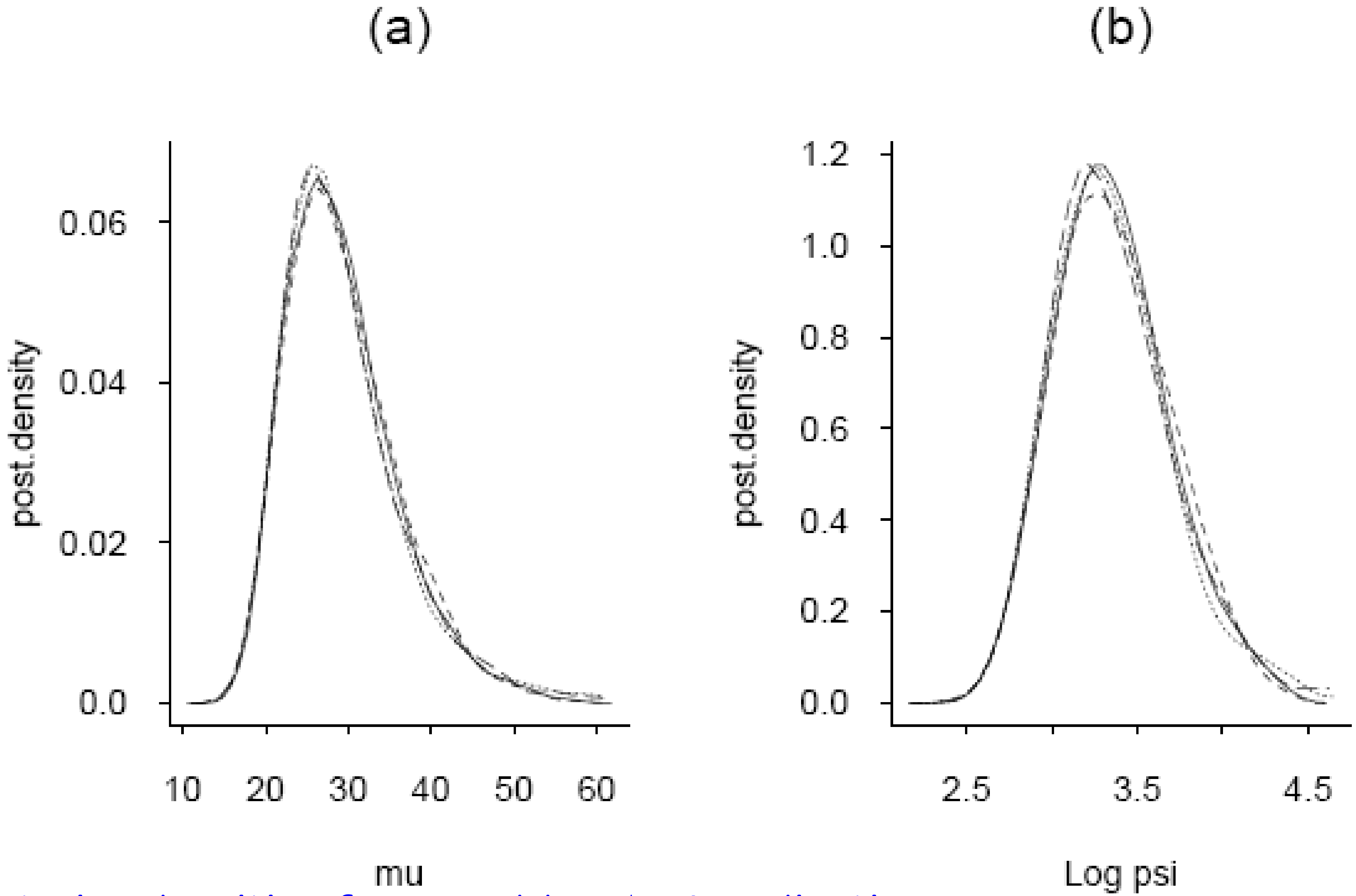
Table 11. Fit of the homogenous point process model to exceedances over various thresholds u. Standard errors are in parentheses.

Fitted parameters for point process model

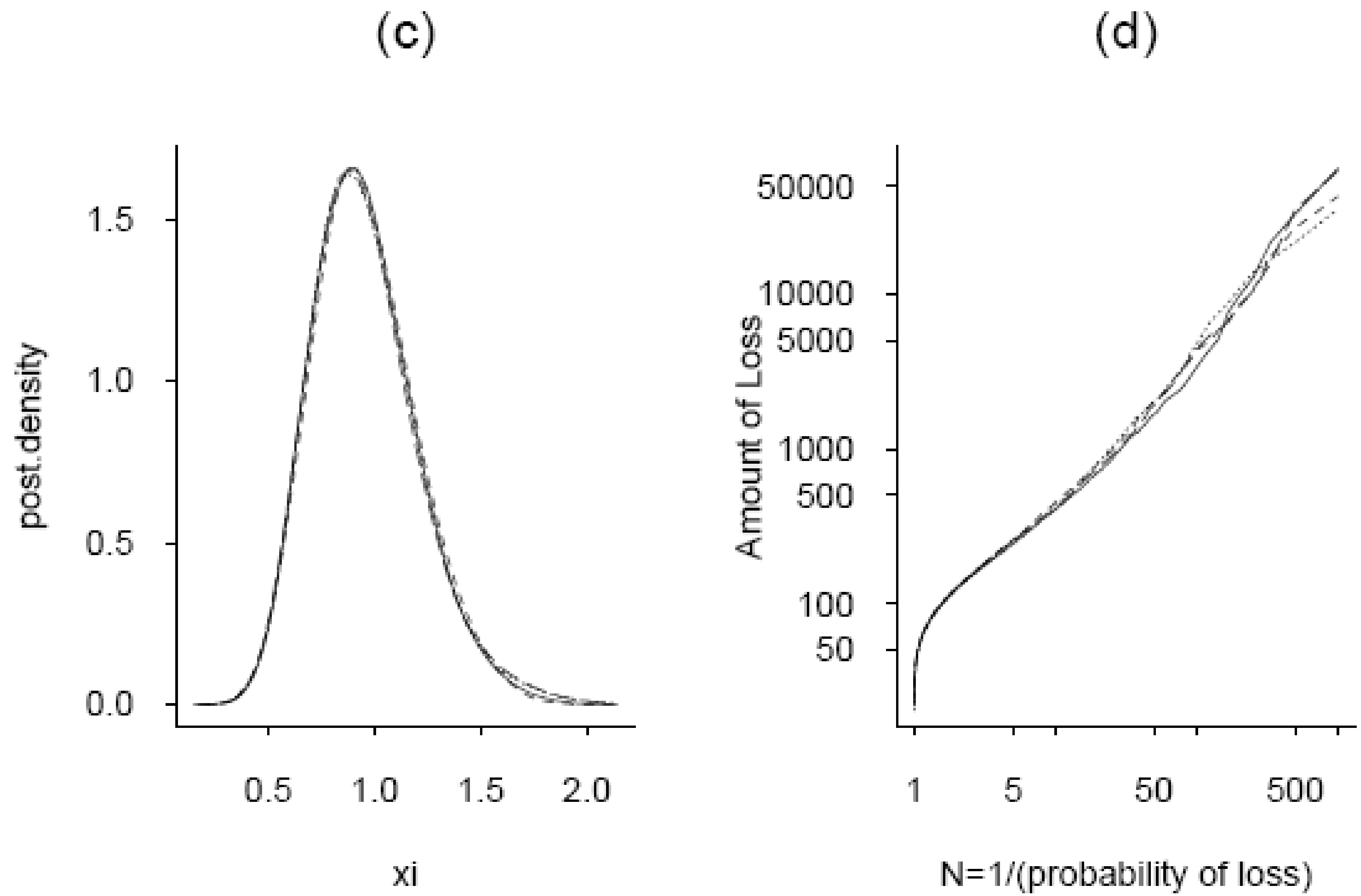


Diagnostic plots for oil company data: $u = 5$

We used a Bayesian analysis to calculate predictive distribution of claim in a future year (similar to Wang Junxia analysis)



Posterior densities for μ and $\log\psi$: 4 realizations



Posterior densities for ξ and predictive DF: 4 realizations

Hierarchical Analysis

Preliminary analyses indicate:

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

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

Level I. Parameters $m_\mu,\ m_\psi,\ m_\xi, s^2_\mu,\ s^2_\psi,\ s^2_\xi$ 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 $N(m_\mu, s^2_\mu)$, the normal distribution with mean m_μ, variance s^2_μ. Similarly, $\log\psi_1, ..., \log\psi_J$ are drawn independently from $N(m_\psi, s^2_\psi)$, $\xi_1, ..., \xi_J$ are drawn independently from $N(m_\xi, s^2_\xi)$.

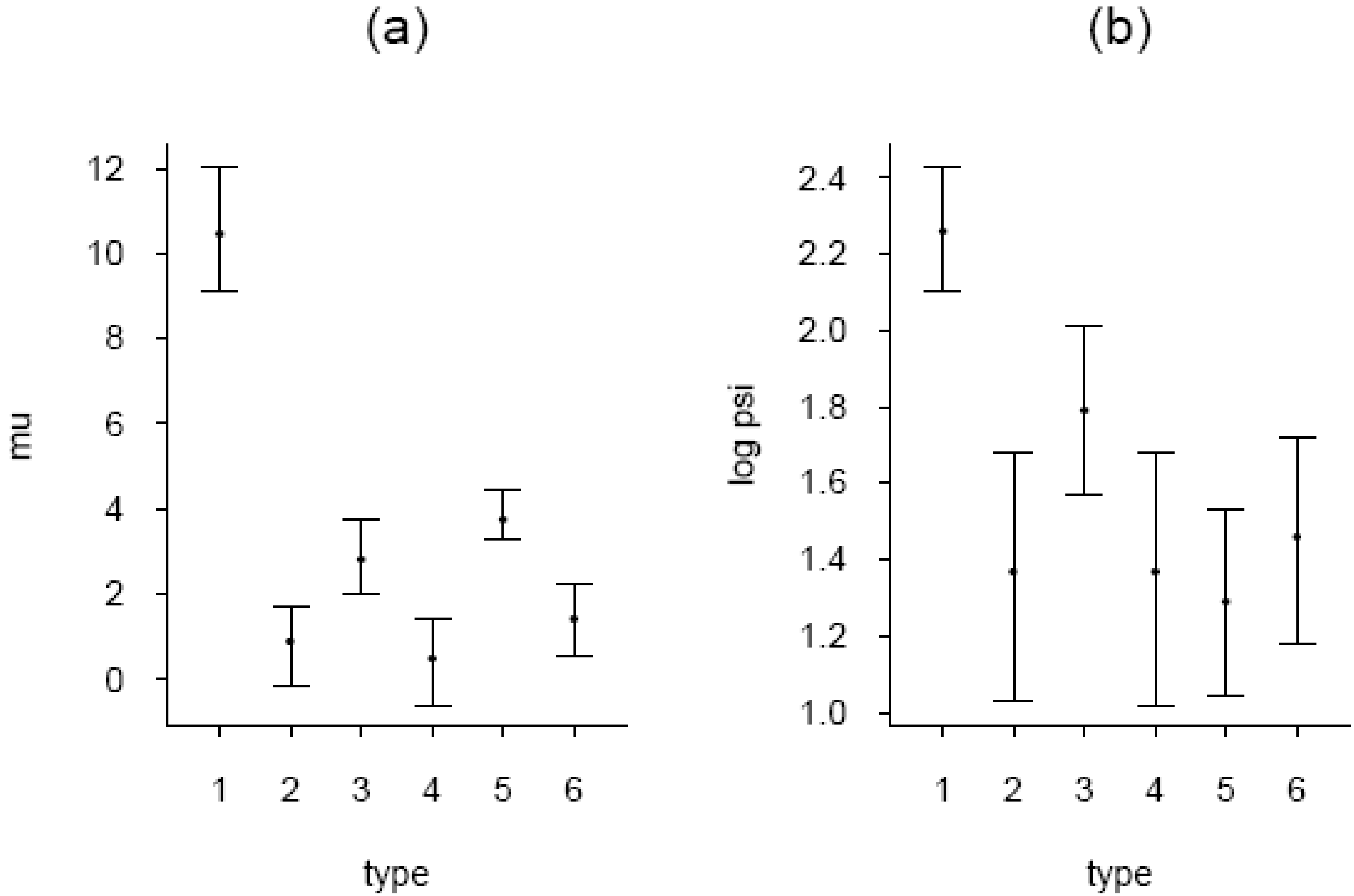
Level III. Conditional on Level II, for each $j \in \{1, ..., J\}$, the point process of exceedances of type j is a realisation from the homogeneous point process model with parameters $\mu_j,\ \psi_j,\ \xi_j$.

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

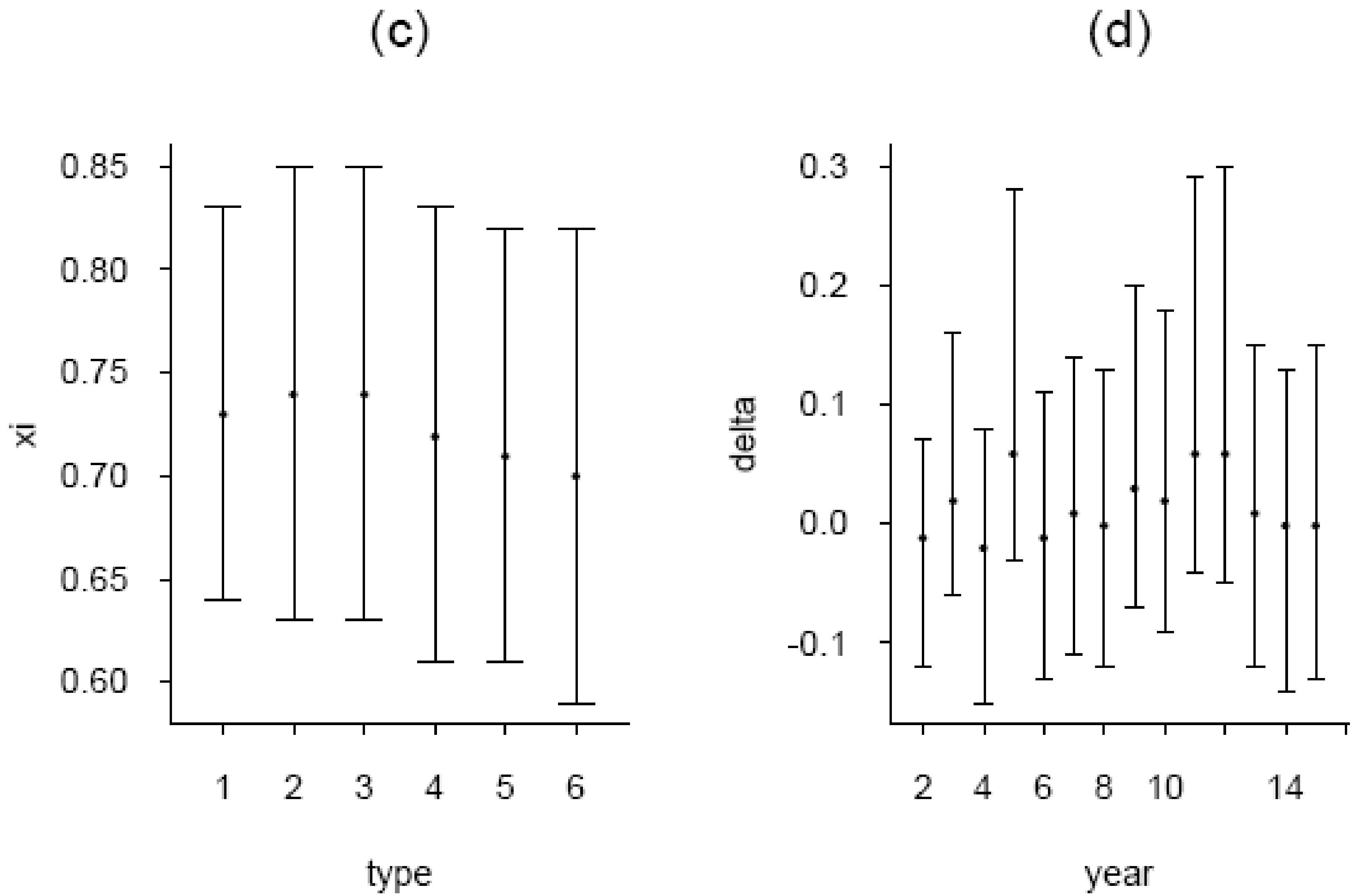
$$\delta_k = \rho\delta_{k-1} + \eta_k, \quad \eta_k \sim N(0, s_\eta^2)$$

with a vague prior on (ρ, s_η^2).

Run hierarchical model by MCMC.



Boxplots for distribution by type of claim



Boxplots for distribution by type of claim and year

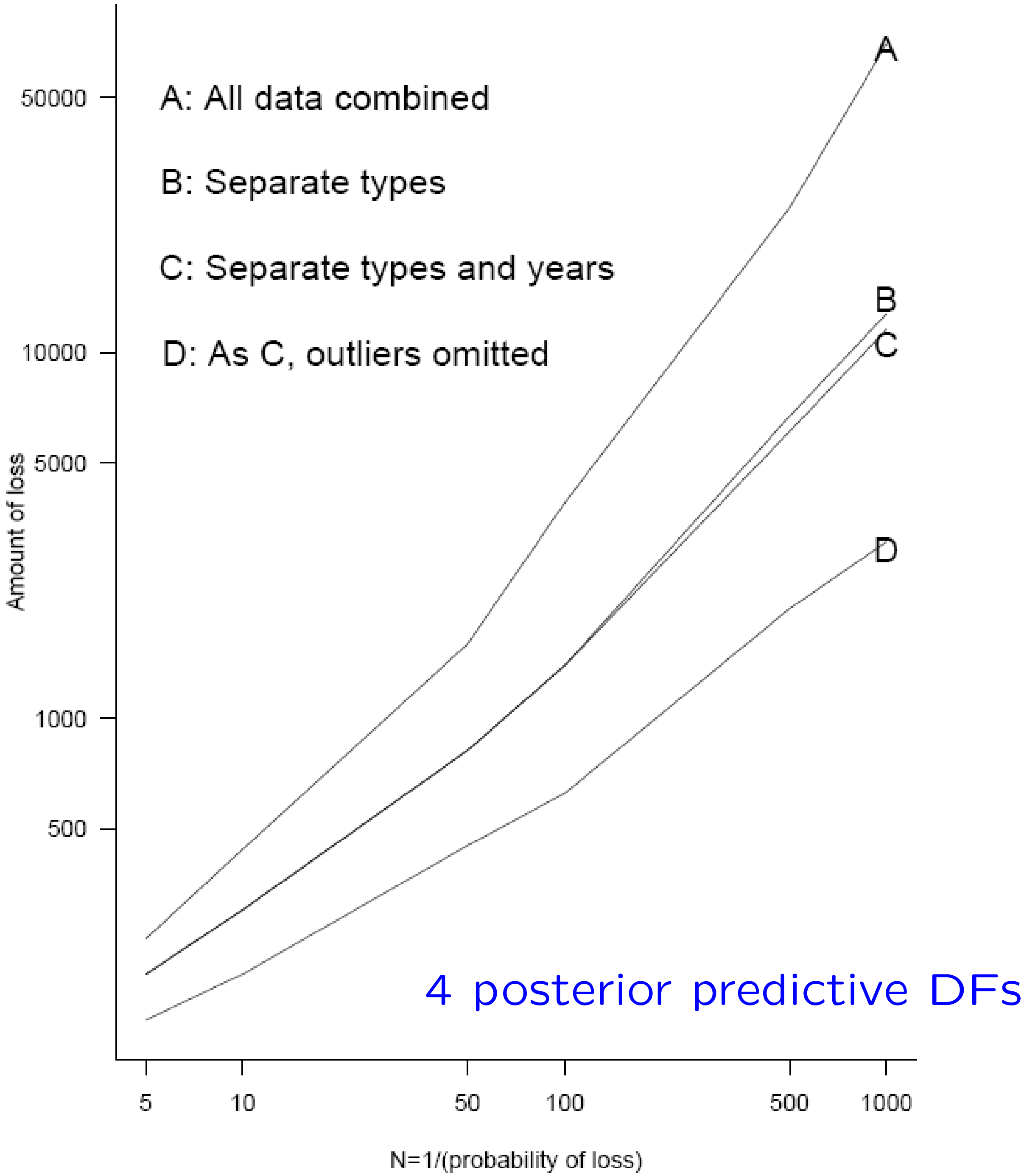
Finally, we collect output from hierarchical model to obtain predictive distribution of future loss in four cases:

A: All data combined into one sample, no identification of outliers, separate type or separate years

B: Identify separate types but not separate years

C: Identify separate types and separate years

D: As C, but omit the two largest observations which may be outliers


A: All data combined
B: Separate types
C: Separate types and years
D: As C, outliers omitted
A
B
C
D
50000
10000
5000
1000
500
Amount of loss
5
10
50
100
500
1000
N=1/(probability of loss)
4 posterior predictive DFs

Take-Home Message

As in the running records example, the Bayesian analysis including a posterior distribution seems the best way to express our uncertainty about future losses.

However, there is an additional benefit in using a hierarchical model. In this case, fitting separate distributions to each category of loss results in less extreme predictions of future losses. This is possibly because the individual categories have less variability than if we combine all the claims into a single sample without regard for category of loss.

VI. 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, *JASA* 2007; Sang and Gelfand, *Environmental and Ecological Statistics*, 2008)

This discussion uses extreme value methodology to look for trends

CURRENT METHODOLOGY
(see e.g. Groisman *et al.* 2005)

Most common method is based on counting exceedances over a high threshold (e.g. 99.7% threshold)

- Express exceedance counts as anomalies from 30-year mean at each station
- Average over regions using "geometric weighting rule": first average within 1° grid boxes, then average grid boxes within region
- Calculate standard error of this procedure using exponential spatial covariances with nugget (range of 30–500 km, nugget-sill ratio of 0–0.7).
- Increasing trends found in many parts of the world, nearly always stronger than trends in precipitation means, but spatially and temporally heterogeneous. Strongest increase in US extreme precipitations is post-1970, about 7% overall

DATA SOURCES

- NCDC Rain Gauge Data
  - 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)

Homogeneous case (cf. Shamseldin *et al.*, preprint):

Consider all exceedances of a high threshold u over time period $[0,T]$ — probability of an exceedance within $(y,y+dy)\times(t,t+dt)$ (for $y>u,\ t\in[0,T]$) of form

$$\frac{1}{\psi}\left(1+\xi\frac{y-\mu}{\psi}\right)_+^{-1/\xi-1}\exp\left\{-\left(1+\xi\frac{u-\mu}{\psi}\right)_+^{-1/\xi}\right\}dydt$$

- $\mu,\ \psi,\ \xi$ are GEV parameters for annual maxima
- N-year return value — the level y_N that is exceeded in any one year with probability $\frac{1}{N}$.

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.

Model selection

Use a sequence of likelihood ratio tests

- For each (K_1, K_2, K_3), construct LRT against some (K_1', K_2', K_3'), $K_1' \geq K_1, K_2' \geq K_2, K_3' \geq K_3$ (not all equal) using standard χ^2 distribution theory

- Look at proportion of rejected tests over all stations. If too high, set $K_j = K_j'$ $(j = 1, 2, 3)$ and repeat procedure

- By trial and error, we select $K_1 = 4, K_2 = 2, K_3 = 1$ (17 model parameters for 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)

Details

- Selection of time-varying threshold — based on the 95th percentile of a 7-day window around the date of interest

- Declustering by r-runs method (Smith and Weissman 1994) — use $r = 1$ for main model runs

- Computation via *sampling the likelihood*: evaluate contributions to likelihood for all observations above threshold, but sample only 5% or 10% of those below, then renormalize to provide accurate approximation to full likelihood

Details (continued)

- Covariances of parameters at different sites:

  $\widehat{\theta}_s$ is MLE at site s, solves $\nabla \ell_s(\widehat{\theta}_s) = 0$

  For two sites s, s',

  $$\text{Cov}\left(\widehat{\theta}_s, \widehat{\theta}_{s'}\right) \approx \left(\nabla^2 \ell_s(\widehat{\theta}_s)\right)^{-1} \text{Cov}\left(\nabla \ell_s(\theta_s), \nabla \ell_{s'}(\theta_{s'})\right) \left(\nabla^2 \ell_{s'}(\widehat{\theta}_{s'})\right)^{-1}$$

  Estimate covariances on RHS empirically, using a subset of days (*same* subset for all stations)

  Also employed when $s = s'$.

  Open question: Should we "regularize" this covariance matrix?

Details (continued)

- Calculating the N-year return value

  For one year ($t = 1, ..., T$), find $y_{\theta,N}$ numerically to solve

$$\sum_{1}^{T} \left(1 + \xi_t \frac{y_{\theta,N} - \mu_t}{\psi_t}\right)^{-1/\xi_t} = -\log\left(1 - \frac{1}{N}\right).$$

  – Also calculate $\frac{\partial y_{\theta,N}}{\partial \theta_j}$ by numerical implementation of inverse function formula

  – Covariances between return level estimates at different sites by

$$\mathrm{Cov}\left\{y_{\widehat{\theta}_s,N}, y_{\widehat{\theta}_{s'},N}\right\} \approx \left(\frac{dy_{\widehat{\theta}_s,N}}{d\theta_s}\right)^T \mathrm{Cov}\left(\widehat{\theta}_s, \widehat{\theta}_{s'}\right)\left(\frac{dy_{\widehat{\theta}_{s'},N}}{d\theta_{s'}}\right).$$

  – Also apply to ratios of return level estimates, such as

$$\frac{25 - \text{year return level at } s \text{ in } 1999}{25 - \text{year return level at } s \text{ in } 1970}$$

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

Details

- Covariates
  - Always include intercept
  - Linear and quadratic terms in elevation and log of mean annual precipitation
  - Contrast "climate space" approach of Cooley *et al.* (2007)

Details (continued)

- Spatial covariances
  - Matérn
  - Exponential with nugget
  - Intrinsically stationary model

$$\mathrm{Var}(Z_s - Z_{s'}) \;=\; \phi_1 d_{s,s'}^{\phi_2} + \phi_3$$

  - Matérn with nugget

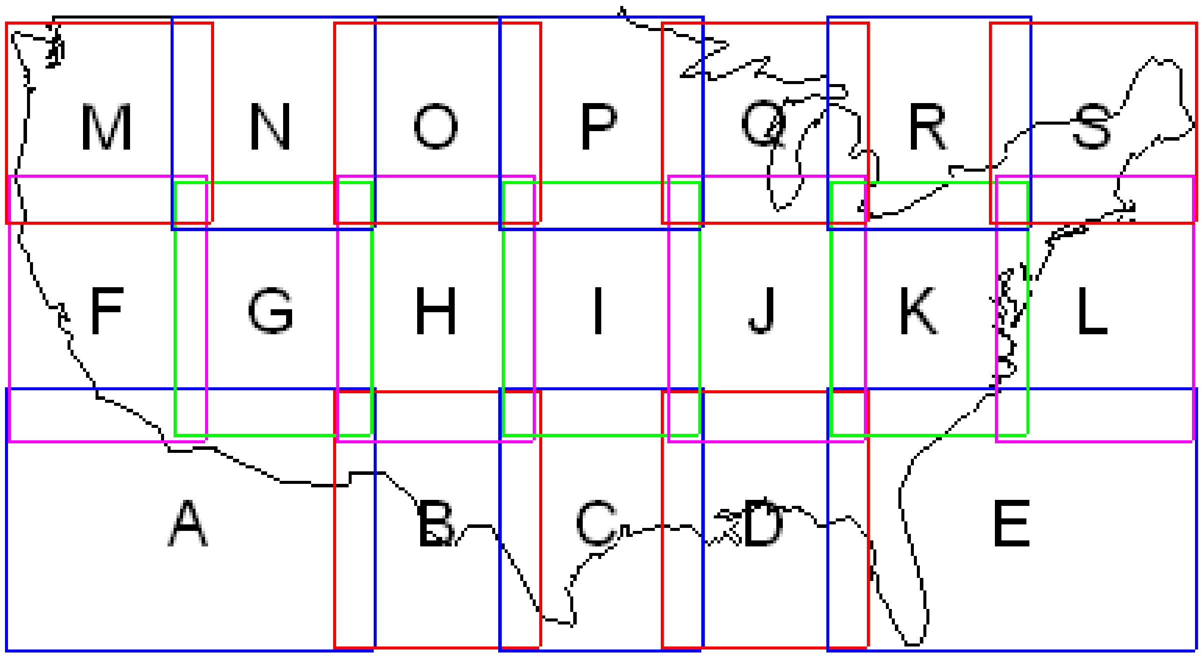
The last-named contains all the previous three as limiting cases and appears to be the best overall, though is often slow to converge (e.g. sometime the range parameter tends to ∞, which is almost equivalent to the intrinsically stationary model)

Details (continued)

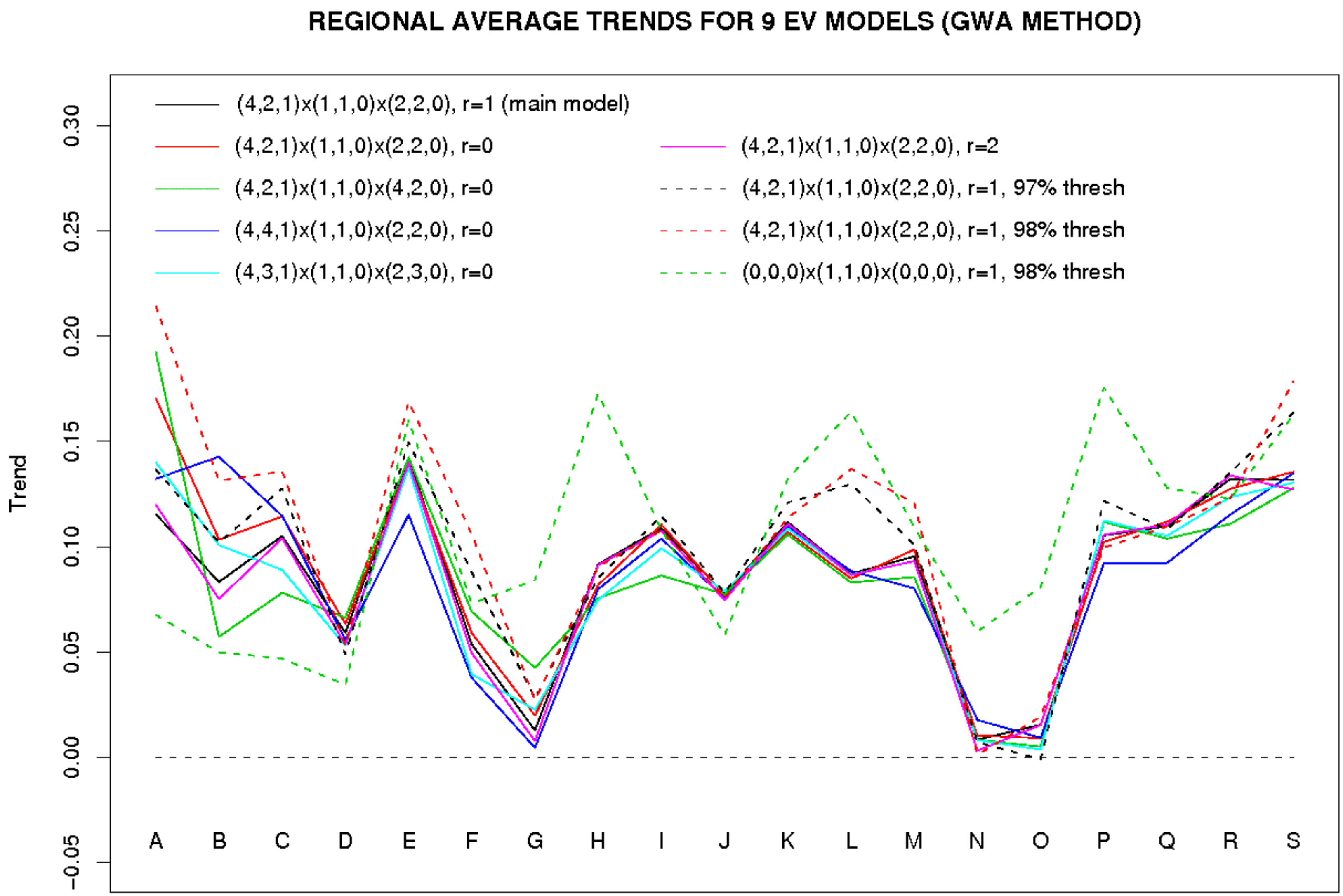
- Spatial heterogeneity

  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.

Summary of models shown on previous slide:

1: Preferred covariates model ($r = 0$ for declustering, uses 95% threshold calculated from 7-day window)

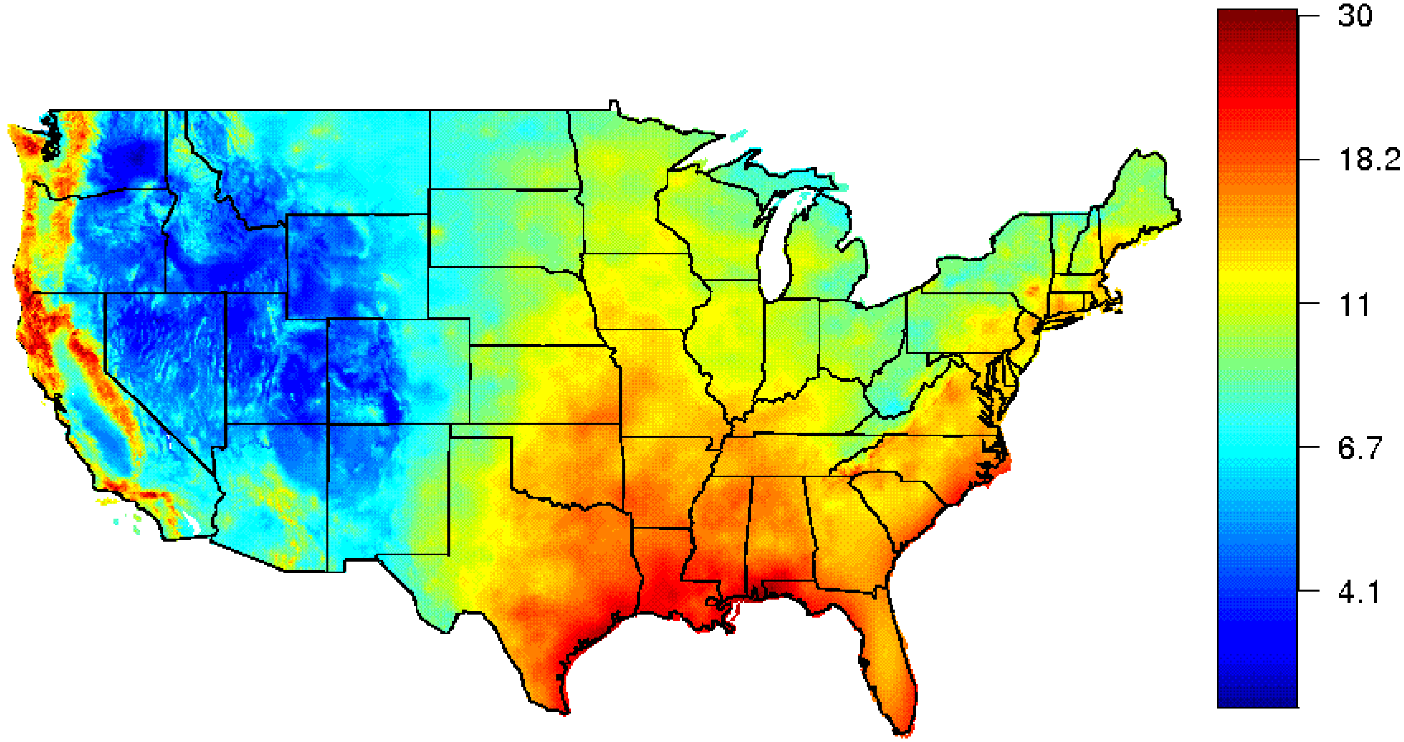
2–4: Three variants where we add covariates to μ_t and/or $\log \psi_t$

5: Replace $r = 0$ by $r = 1$ (subsequent results are based on this model)

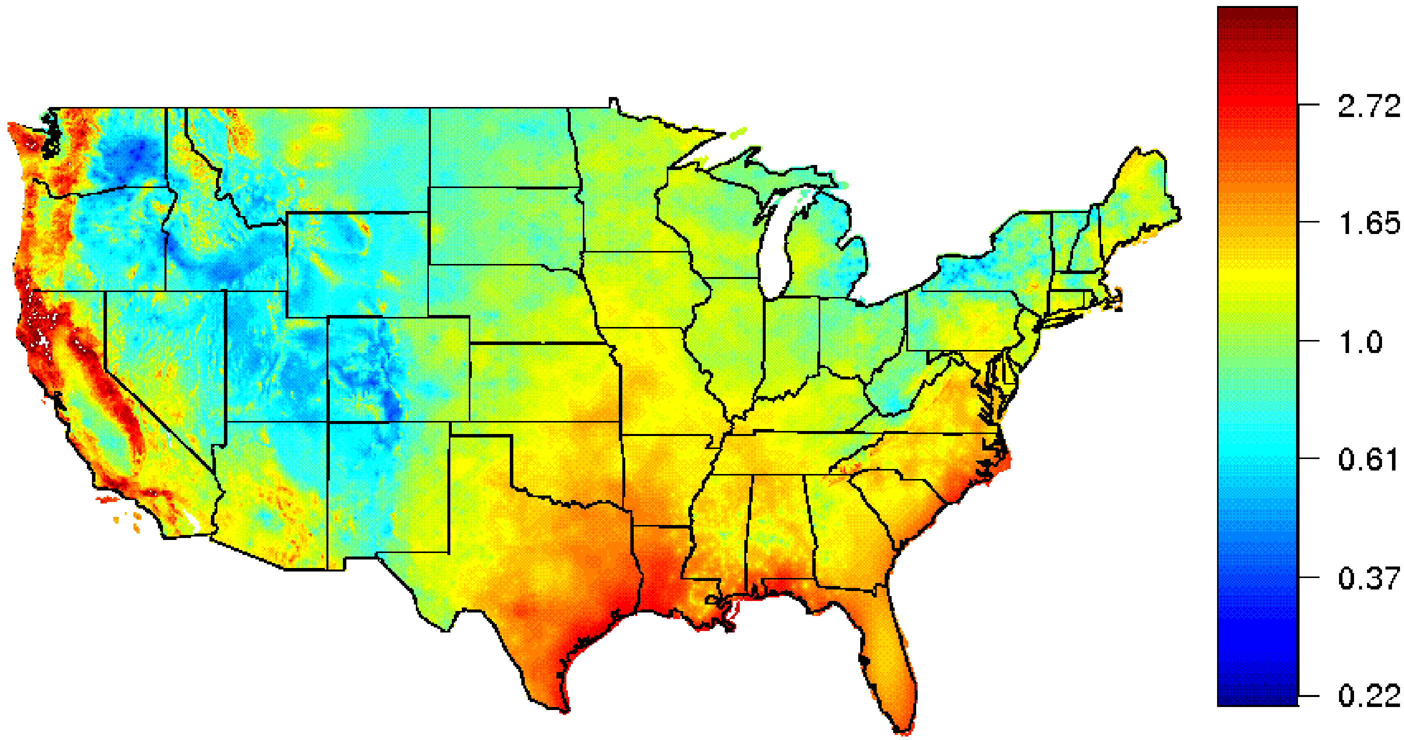
6: Replace $r = 0$ by $r = 2$

7: 97% threshold calculated from 14-day window

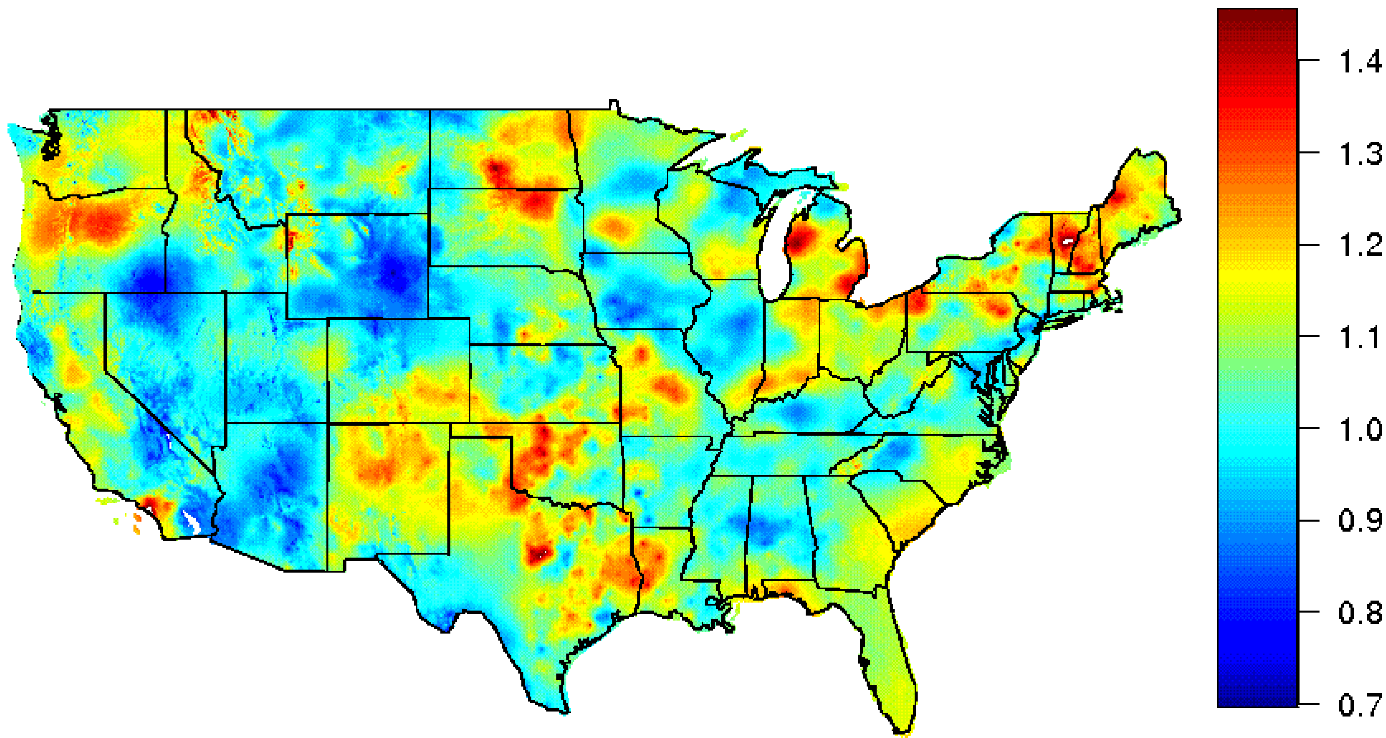
8: 98% threshold calculated from 28-day window



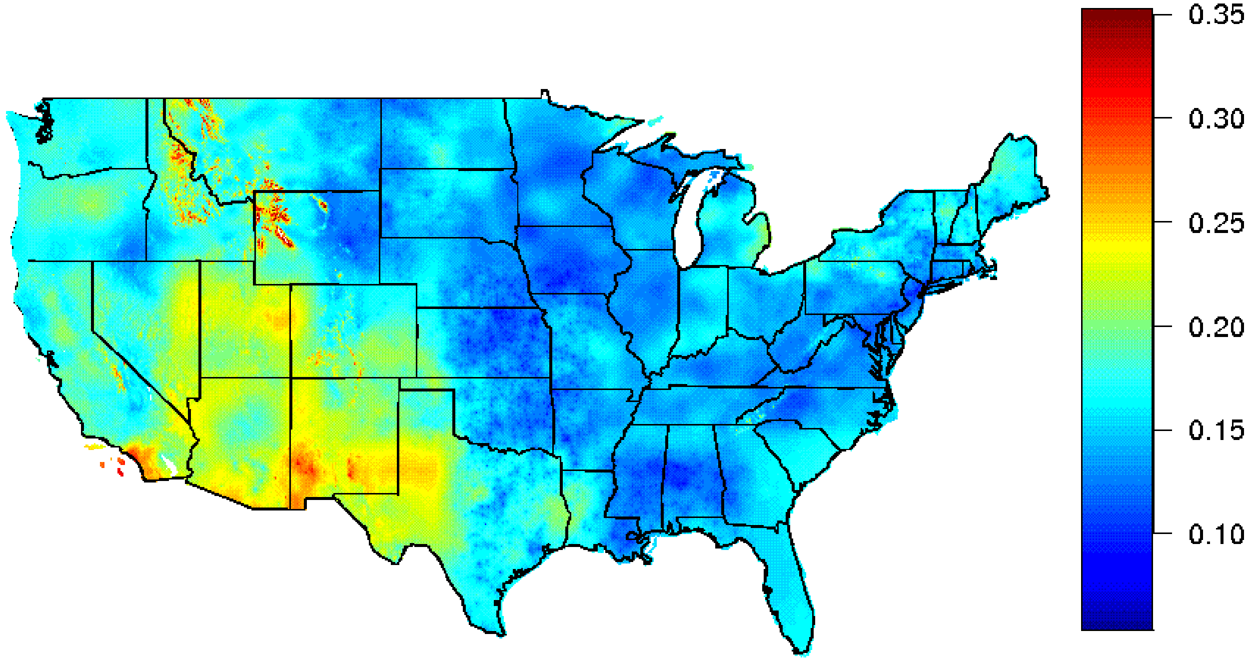
Map of 25-year return values (cm.) for the years 1970–1999



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



Root mean square prediction errors for map of ratios of 25-year return values in 1999 to those in 1970

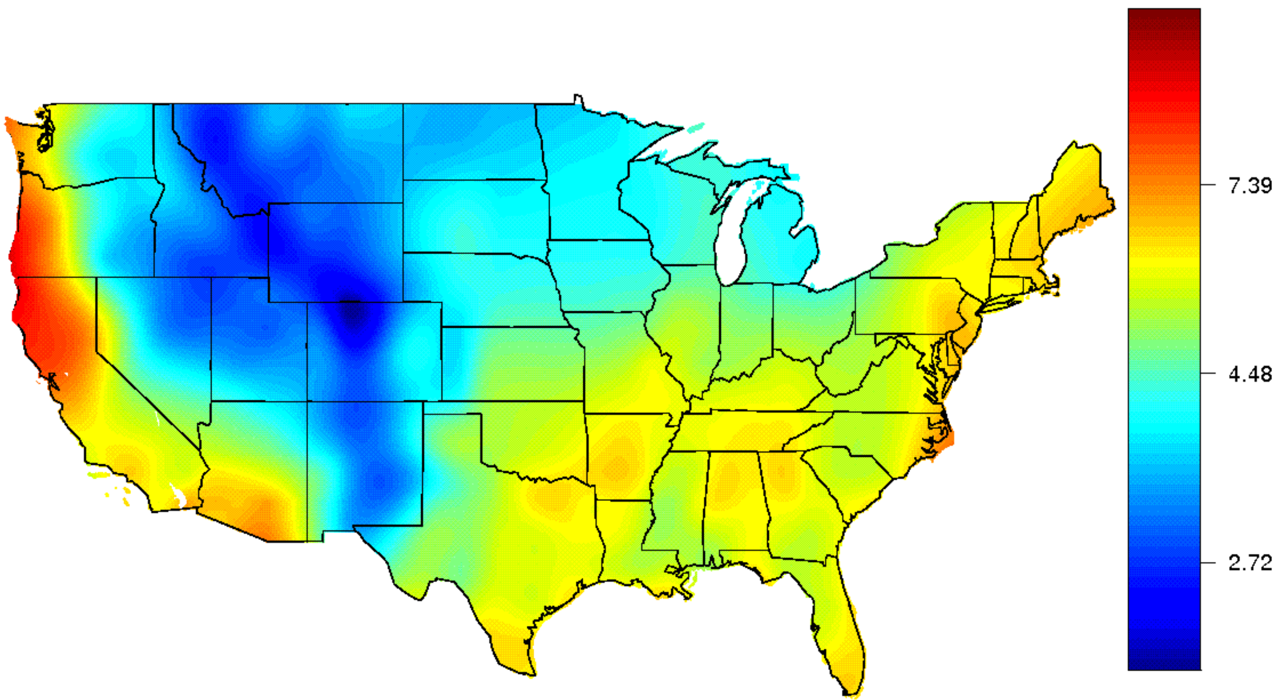
	Δ_1	S_1	Δ_2	S_2		Δ_1	S_1	Δ_2	S_2
A	−0.01	.03	0.05**	.05	K	0.08***	.01	0.09**	.03
B	0.07**	.03	0.08***	.04	L	0.07***	.02	0.07*	.04
C	0.11***	.01	0.10	.03	M	0.07***	.02	0.10**	.03
D	0.05***	.01	0.06	.05	N	0.02	.03	0.01	.03
E	0.13***	.02	0.14*	.05	O	0.01	.02	0.02	.03
F	0.00	.02	0.05*	.04	P	0.07***	.01	0.11***	.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

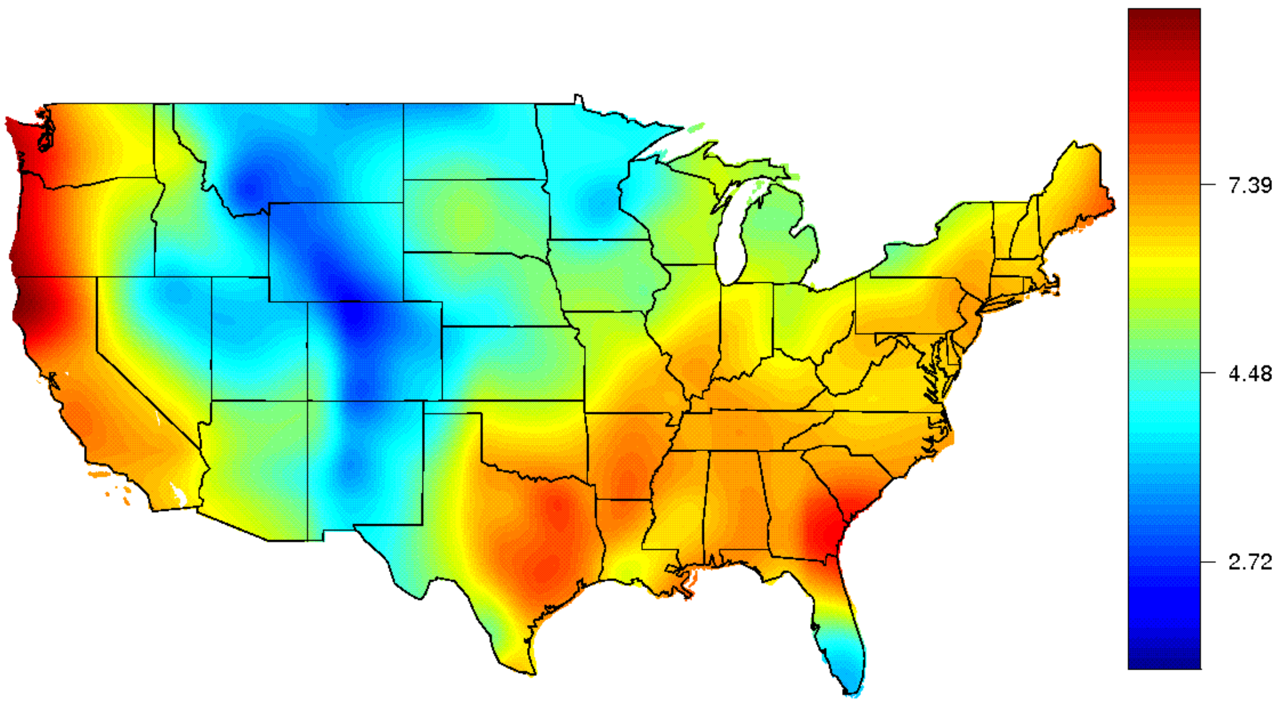
S_1: Corresponding standard error (or RMSPE)

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

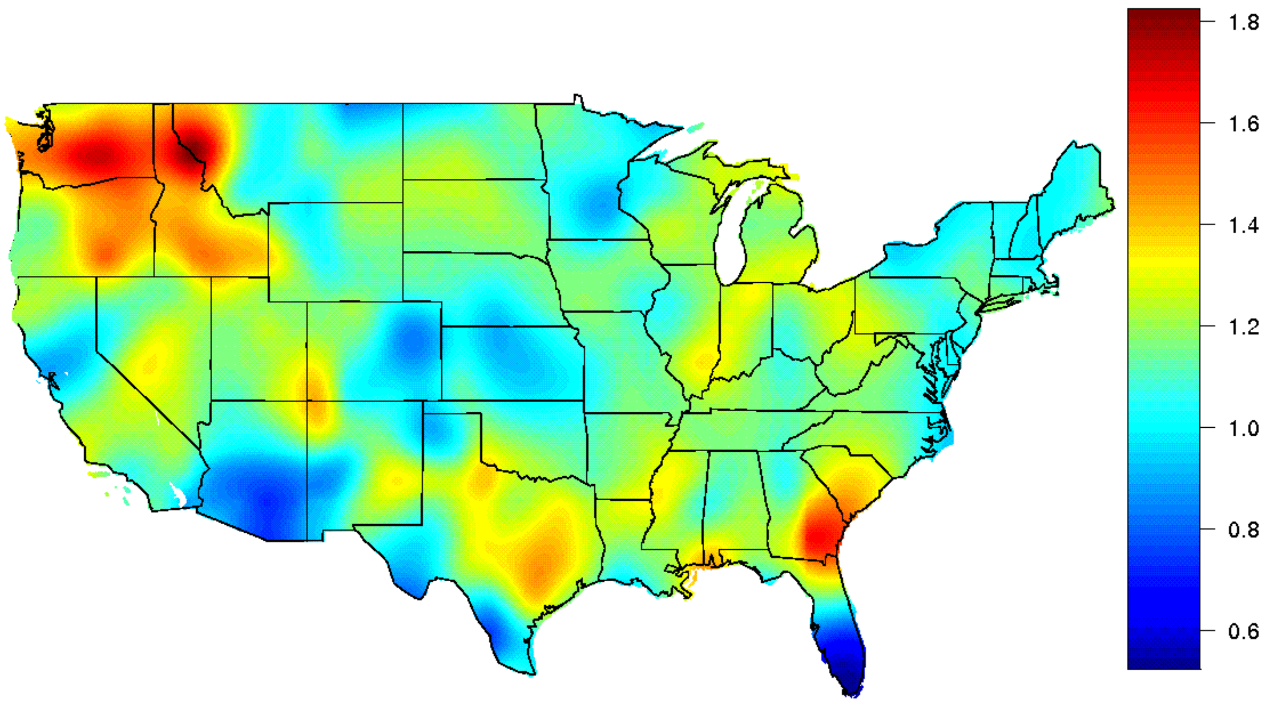
Stars indicate significance at 5%*, 1%**, 0.1%***.



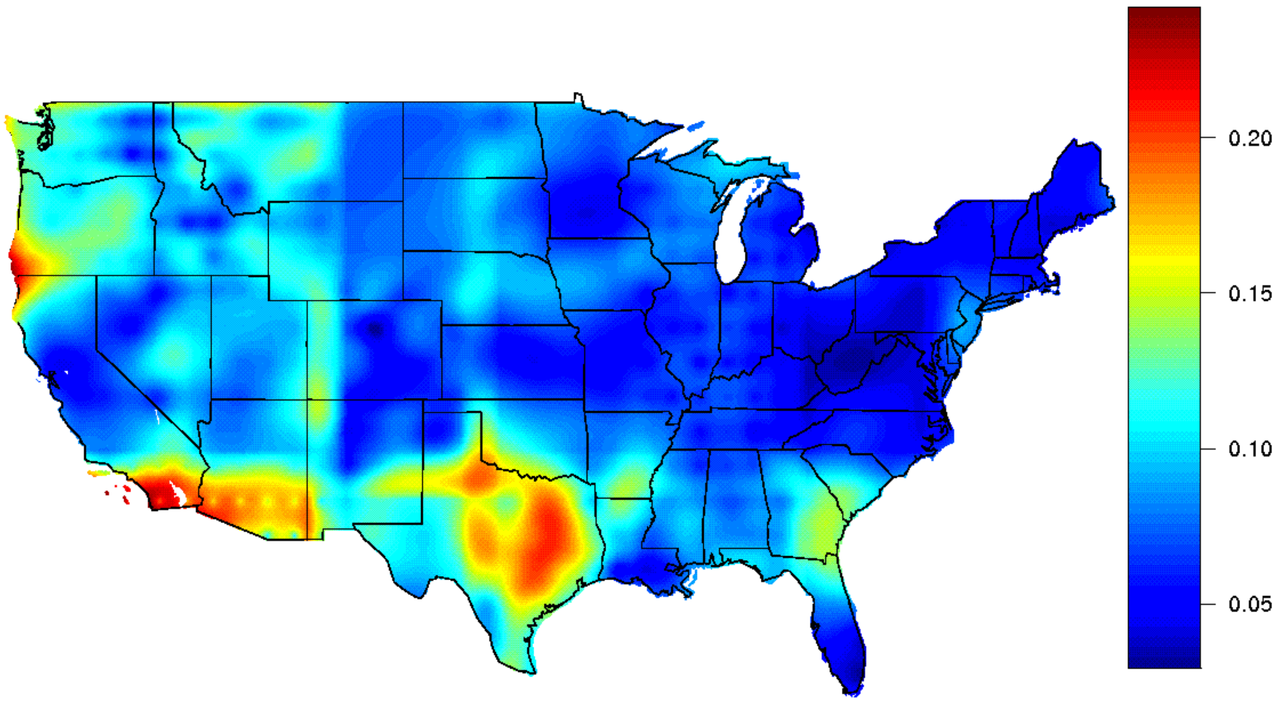
Return value map for CCSM data (cm.): 1970–1999



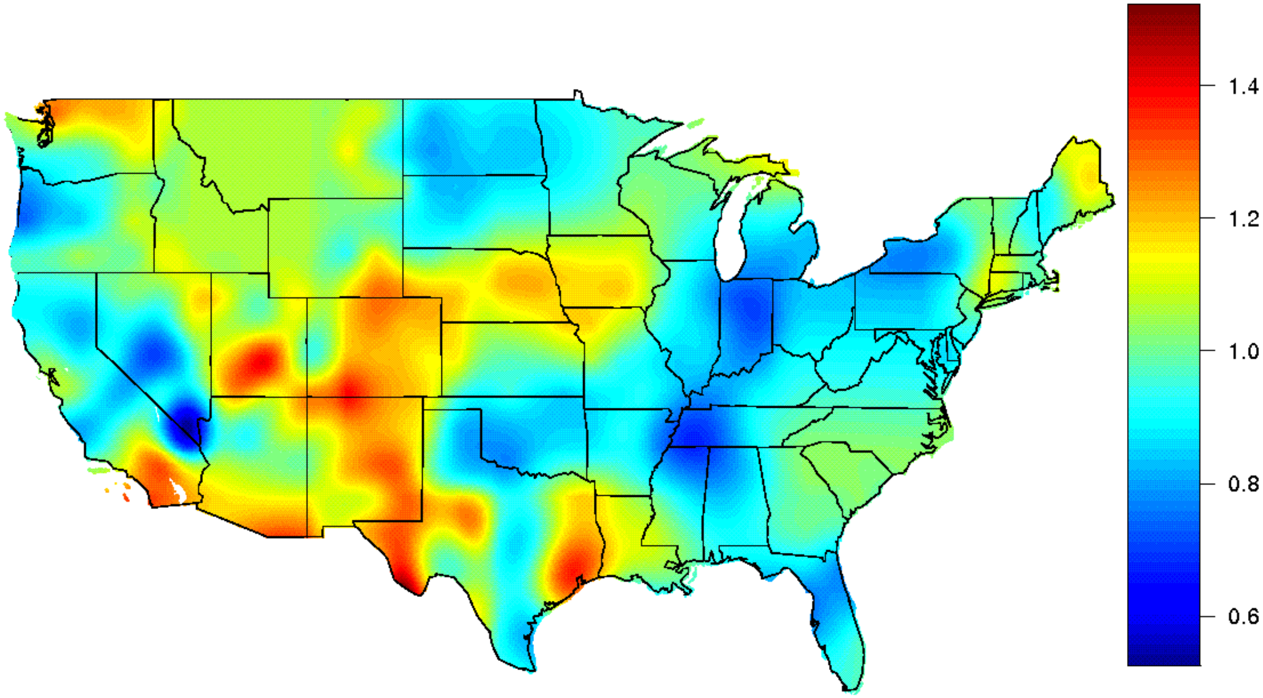
Return value map for CCSM data (cm.): 2070–2099



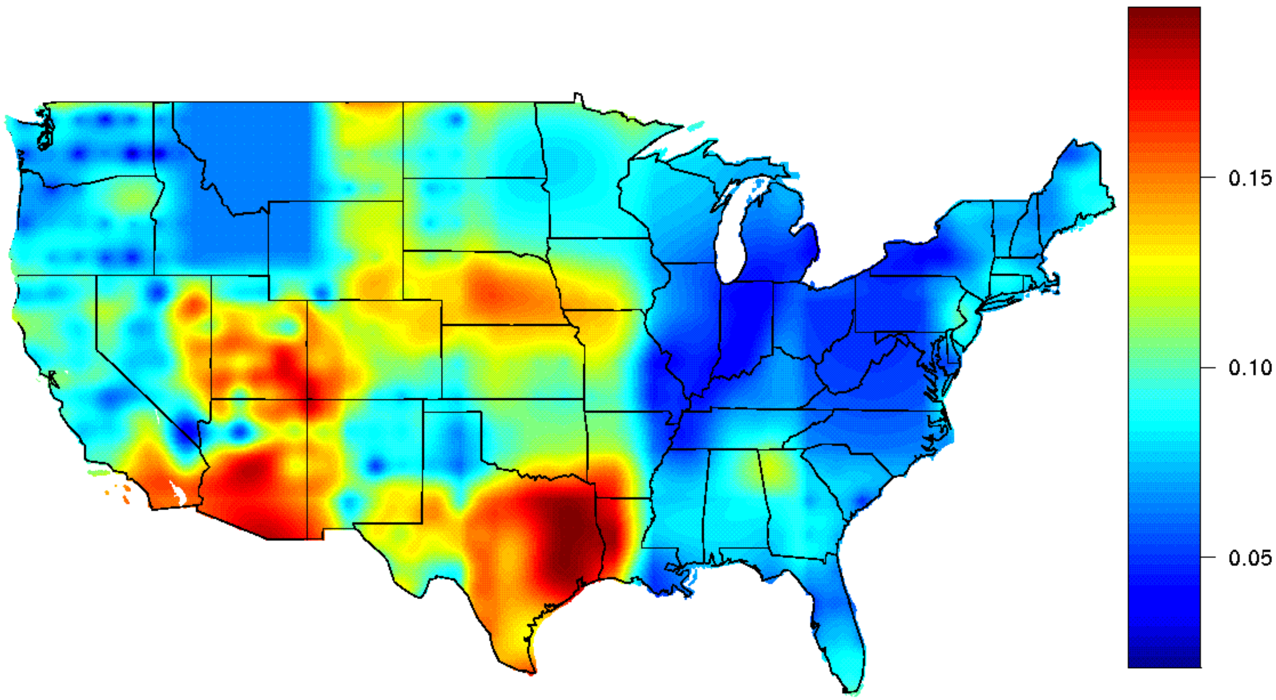
Estimated ratios of 25-year return values for 2070–2099 to those of 1970–1999, based on CCSM data, A2 scenario



RMSPE for map in previous slide



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



RMSPE for map in previous slide

	Δ_3	S_3	Δ_4	S_4		Δ_3	S_3	Δ_4	S_4
A	0.16^{**}	.07	0.24^{**}	.10	K	-0.08^{***}	.02	-0.11^{*}	.05
B	0.14^{***}	.04	0.12^{***}	.06	L	-0.04	.04	-0.03	.06
C	0.02	.05	-0.14	.11	M	0.01	.03	0.00	.08
D	-0.06	.04	-0.15	.10	N	0.06^{**}	.02	0.05	.06
E	-0.07^{*}	.03	-0.09	.08	O	-0.03	.04	-0.06	.07
F	-0.07^{*}	.04	-0.03	.05	P	-0.01	.04	-0.07	.07
G	0.03	.03	0.08^{*}	.04	Q	-0.04	.04	-0.03	.07
H	0.11^{***}	.03	0.08	.06	R	-0.17^{***}	.03	-0.06	.08
I	-0.02	.04	-0.05	.07	S	0.00	.04	0.02	.05
J	-0.15^{***}	.03	-0.16^{**}	.06					

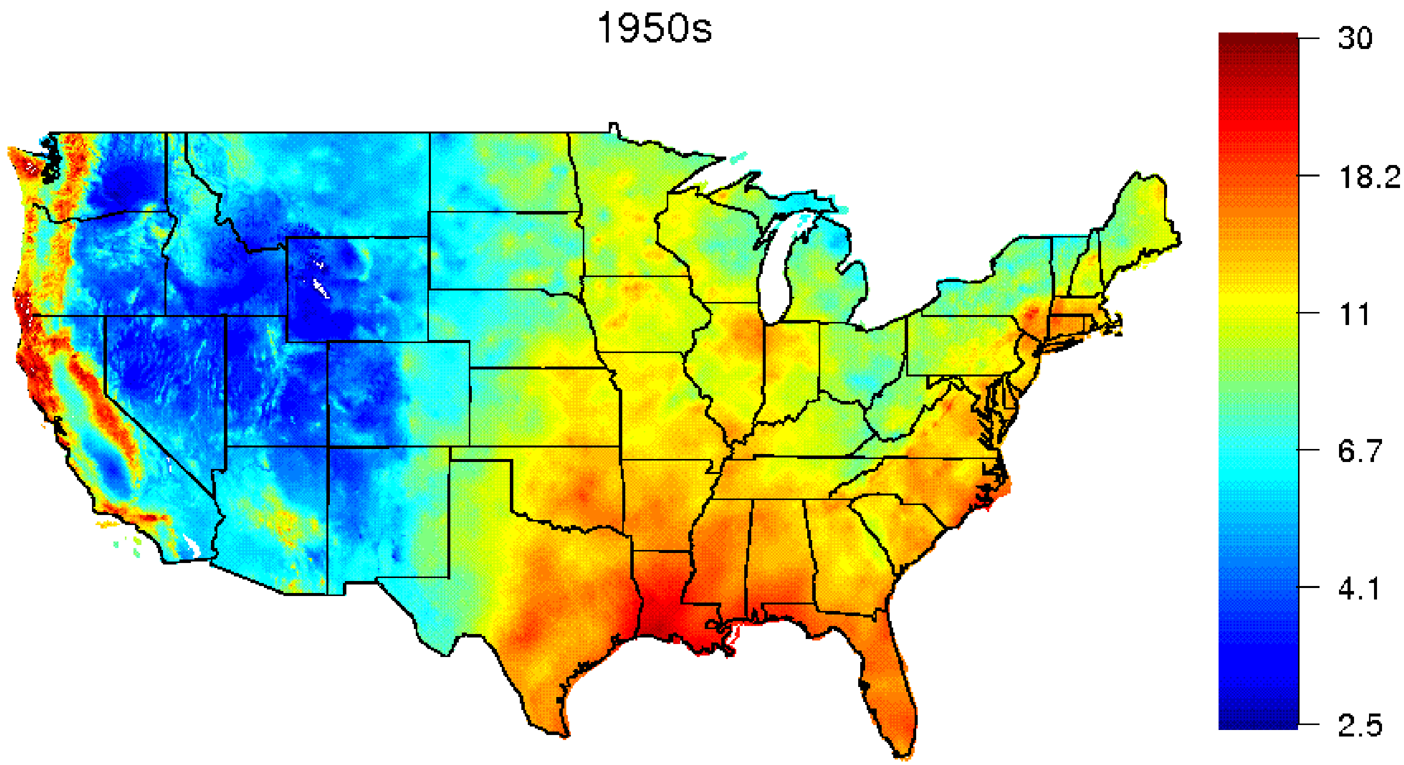
Δ_3: Mean change in log 25-year return value (1970 to 1999) for CCSM, by kriging

SE_3: Corresponding standard error (or RMSPE)

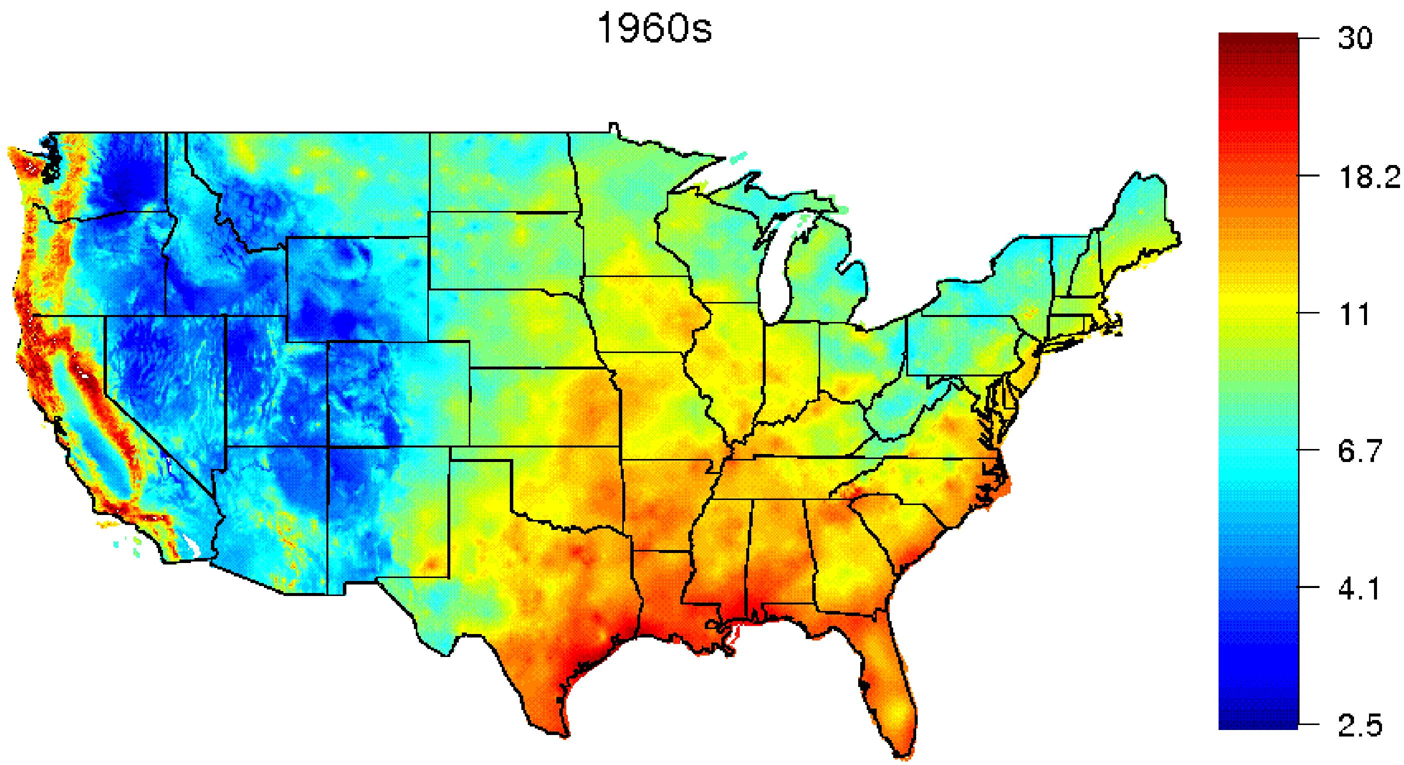
Δ_4, SE_4: Results using GWA

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

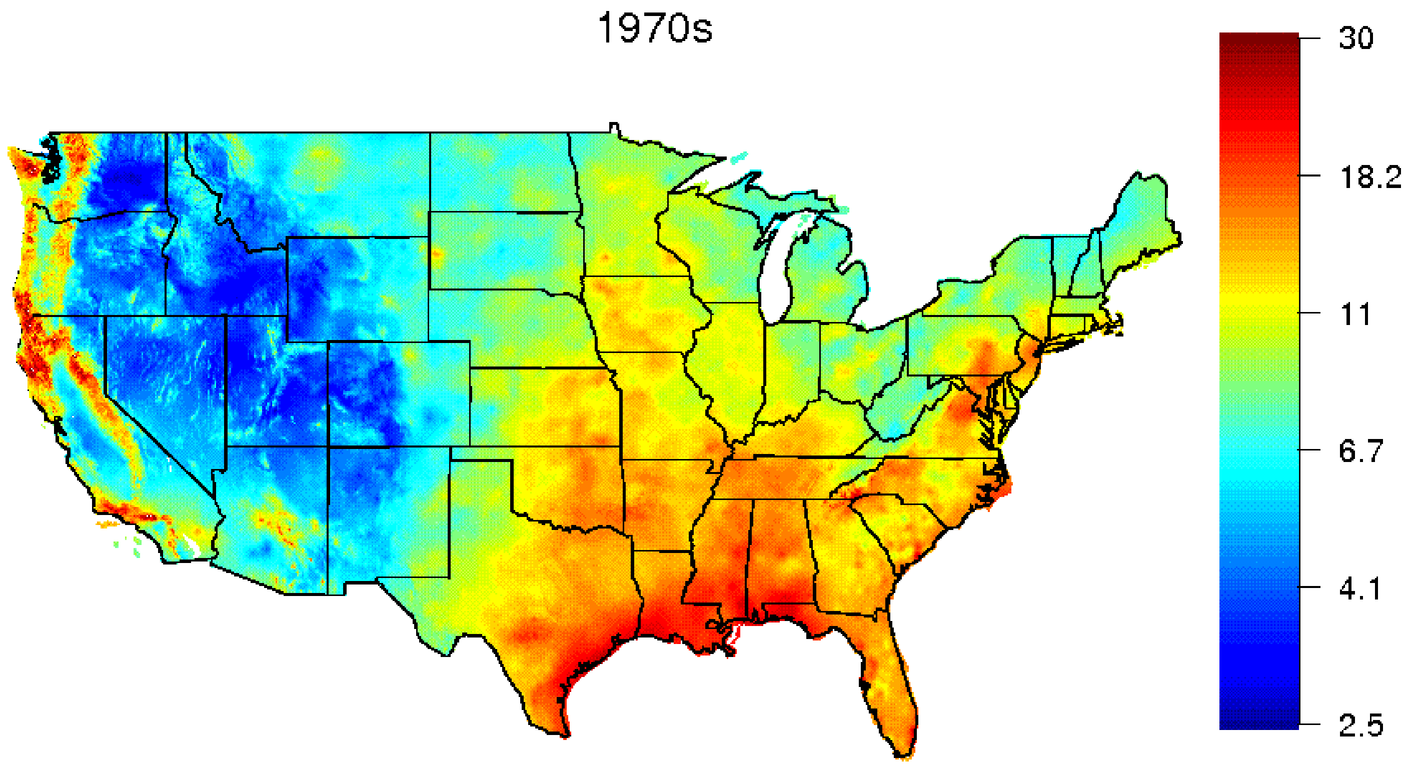
RETURN VALUE MAPS FOR INDIVIDUAL DECADES



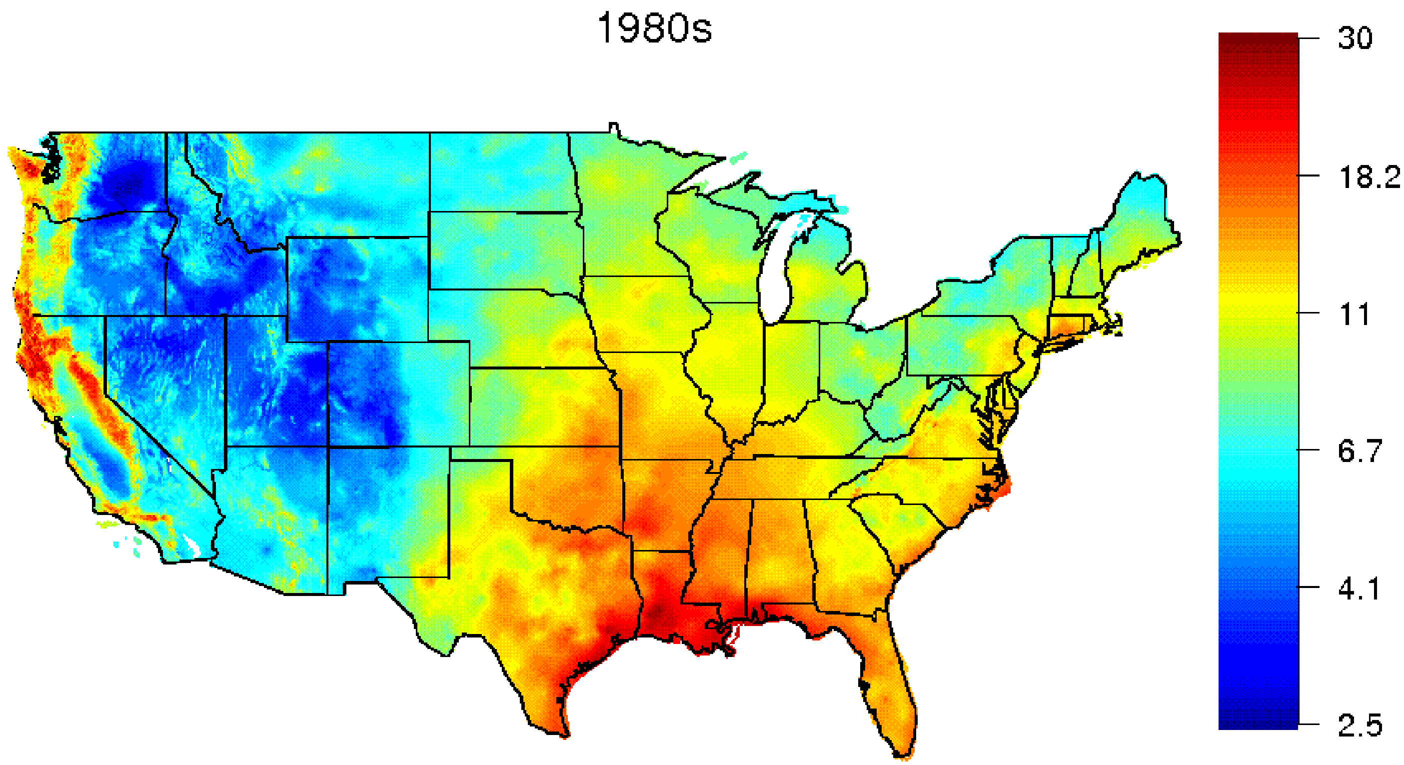
Map of 25-year return values (cm.) for the years 1950–1959



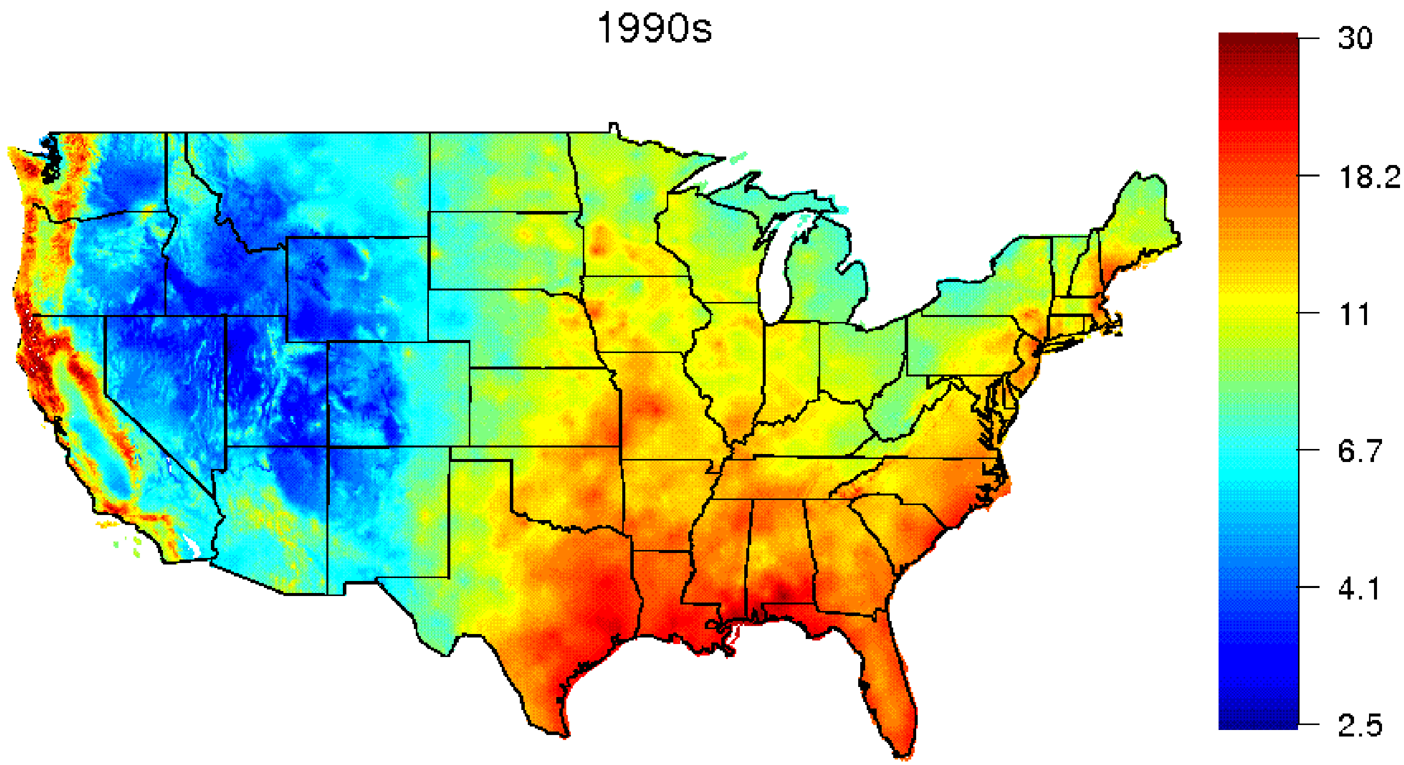
Map of 25-year return values (cm.) for the years 1960–1969



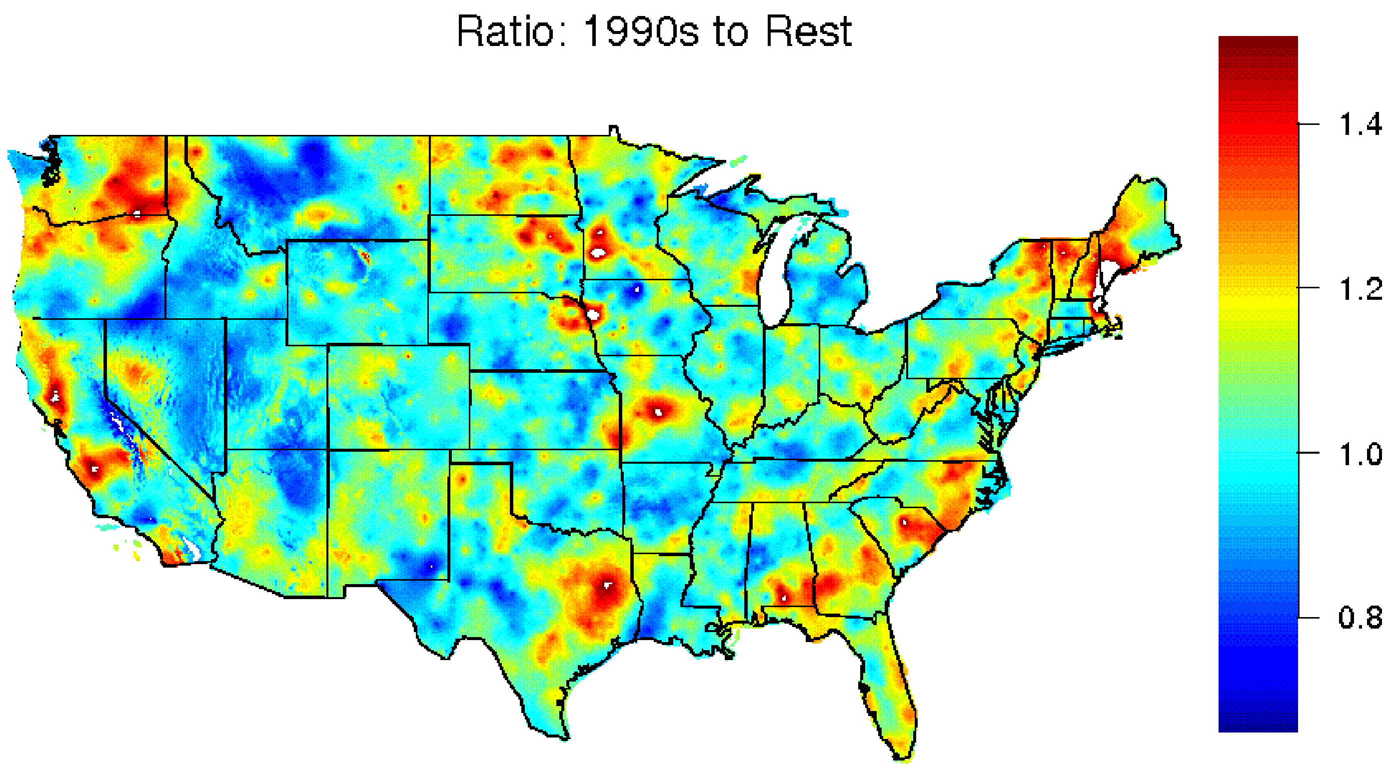
Map of 25-year return values (cm.) for the years 1970–1979



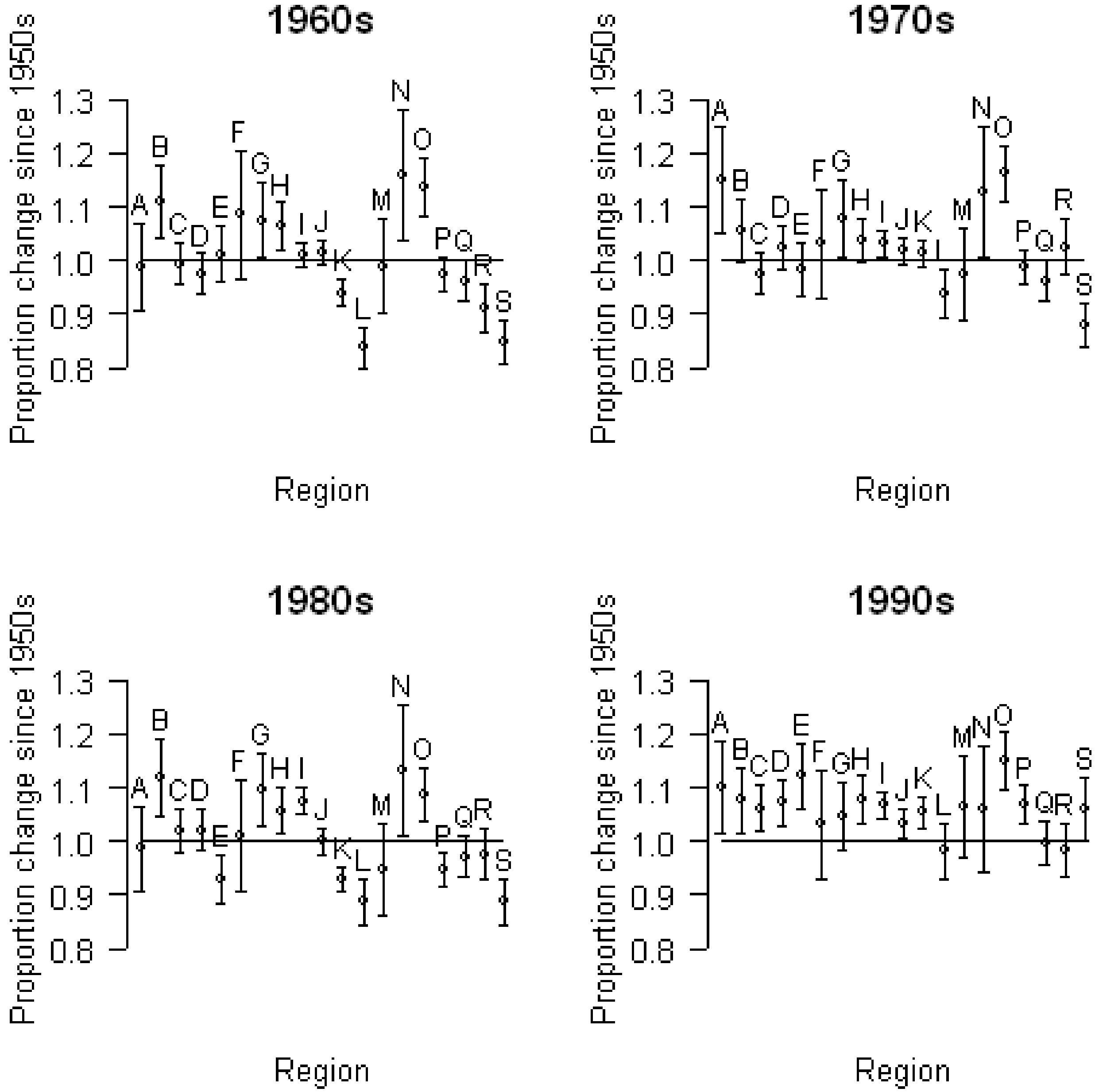
Map of 25-year return values (cm.) for the years 1980–1989



Map of 25-year return values (cm.) for the years 1990–1999



Estimated ratios of 25-year return values for 1990s compared with average at each location over 1950–1989



Regional changes in log RV25 for each decade compared with 1950s

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. Kriging better than GWA

5. *But...* based on CCSM data there is a completely different spatial pattern and no overall increase

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

7. Decadal variations since 1950s show strongest increases during 1990s.

VII. MULTIVARIATE EXTREMES AND MAX-STABLE PROCESSES

VII.1 Limit theory for multivariate sample maxima

VII.2 Alternative formulations of multivariate extreme value theory

VII.3 Max-stable processes

Multivariate extreme value theory applies when we are interested in the joint distribution of extremes from several random variables.

Examples:

- Winds and waves on an offshore structure
- Meteorological variables, e.g. temperature and precipitation
- Air pollution variables, e.g. ozone and sulfur dioxide
- Finance, e.g. price changes in several stocks or indices
- Spatial extremes, e.g. joint distributions of extreme precipitation at several locations

VII.1 LIMIT THEOREMS FOR MULTIVARIATE SAMPLE MAXIMA

Let $\mathbf{Y}_i = (Y_{i1}...Y_{iD})^T$ be i.i.d. D-dimensional vectors, $i = 1, 2, ...$

$M_{nd} =$ max$\{Y_{1d}, ..., Y_{nd}\}(1 \le d \le D)$ — d'th-component maximum

Look for constants a_{nd}, b_{nd} such that

$$\Pr\left\{\frac{M_{nd} - b_{nd}}{a_{nd}} \le x_d,\ d = 1, ..., D\right\} \ \rightarrow \ G(x_1, ..., x_D).$$

Vector notation:

$$\Pr\left\{\frac{\mathbf{M}_n - \mathbf{b}_n}{\mathbf{a}_n} \le \mathbf{x}\right\} \ \rightarrow \ G(\mathbf{x}).$$

Two general points:

1. If G is a multivariate extreme value distribution, then each of its marginal distributions must be one of the univariate extreme value distributions, and therefore can be represented in GEV form

2. The form of the limiting distribution is invariant under monotonic transformation of each component. Therefore, without loss of generality we can transform each marginal distribution into a specified form. For most of our applications, it is convenient to assume the Fréchet form:

$$\Pr\{X_d \le x\} \;=\; \exp(-x^{-\alpha}),\ x > 0,\ d = 1, ..., D.$$

Here $\alpha > 0$. The case $\alpha = 1$ is called *unit Fréchet.*

Representations of Limit Distributions

Any MEVD with unit Fréchet margins may be written as

$$G(\mathbf{x}) = \exp\{-V(\mathbf{x})\}, \tag{1}$$

$$V(\mathbf{x}) = \int_{\mathcal{S}_D} \max_{d=1,\ldots,D} \left(\frac{w_j}{x_j}\right) dH(w), \tag{2}$$

where $\mathcal{S}_D = \{(x_1, \ldots, x_D) : \; x_1 \geq 0, \ldots, x_D \geq 0, x_1 + \ldots + x_D = 1\}$ and H is a measure on $\mathcal{S}_D$.

The function $V(\mathbf{x})$ is called the *exponent measure* and formula (2) is the *Pickands representation*. If we fix $d' \in \{1, \ldots, D\}$ with $0 < x_{d'} < \infty$, and define $x_d = +\infty$ for $d \neq d'$, then

$$V(\mathbf{x}) = \int_{\mathcal{S}_D} \max_{d=1,\ldots,D} \left(\frac{w_d}{x_d}\right) dH(w) = \frac{1}{x_{d'}} \int_{\mathcal{S}_D} w_{d'} dH(w)$$

so to ensure correct marginal distributions we must have

$$\int_{\mathcal{S}_D} w_d dH(w) = 1, \quad d = 1, \ldots, D. \tag{3}$$

Note that

$$kV(\mathbf{x}) = V\left(\frac{\mathbf{x}}{k}\right)$$

(which is in fact another characterization of V) so

$$\begin{aligned} G^k(\mathbf{x}) &= \exp(-kV(\mathbf{x})) \\ &= \exp\left(-V\left(\frac{\mathbf{x}}{k}\right)\right) \\ &= G\left(\frac{\mathbf{x}}{k}\right). \end{aligned}$$

Hence G is *max-stable*. In particular, if $\mathbf{X}_1, ..., \mathbf{X}_k$ are i.i.d. from G, then max$\{\mathbf{X}_1, ..., \mathbf{X}_k\}$ (vector of componentwise maxima) has the same distribution as $k\mathbf{X}_1$.

The Pickands Dependence Function

In the two-dimensional case (with unit Fréchet margins) there is an alternative representation due to Pickands:

$$\Pr\{X_1 \le x_1, X_2 \le x_2\} = \exp\left\{-\left(\frac{1}{x_1}+\frac{1}{x_2}\right) A\left(\frac{x_1}{x_1+x_2}\right)\right\}$$

where A is a convex function on [0,1] such that $A(0) = A(1) = 1$, $\frac{1}{2} \le A\left(\frac{1}{2}\right) \le 1$.

The value of $A\left(\frac{1}{2}\right)$ is often taken as a measure of the strength of dependence: $A\left(\frac{1}{2}\right) = \frac{1}{2}$ is "perfect dependence" while $A\left(\frac{1}{2}\right) = 1$ is independence.

VII.1.a EXAMPLES

Logistic (Gumbel and Goldstein, 1964)

$$V(\mathbf{x}) = \left(\sum_{d=1}^{D} x_d^{-r} \right)^{1/r}, \quad r \geq 1.$$

Check:

1. $V(\mathbf{x}/k) = kV(\mathbf{x})$
2. $V((+\infty, +\infty, ..., x_d, ..., +\infty, +\infty) = x_d^{-1}$
3. $e^{-V(\mathbf{x})}$ is a valid c.d.f.

Limiting cases:

- $r = 1$: independent components
- $r \to \infty$: the limiting case when $X_{i1} = X_{i2} = ... = X_{iD}$ with probability 1.

Asymmetric logistic (Tawn 1990)

$$V(\mathbf{x}) \;=\; \sum_{c\in C}\left\{\sum_{i\in c}\left(\frac{\theta_{i,c}}{x_i}\right)^{r_c}\right\}^{1/r_c},$$

where C is the class of non-empty subsets of $\{1,\dots,D\}$, $r_c \geq 1$, $\theta_{i,c} = 0$ if $i \notin c$, $\theta_{i,c} \geq 0$, $\sum_{c\in C}\theta_{i,c} = 1$ for each i.

Negative logistic (Joe 1990)

$$V(\mathbf{x}) \;=\; \sum\frac{1}{x_j} + \sum_{c\in C:\ |c|\geq 2}(-1)^{|c|}\left\{\sum_{i\in c}\left(\frac{\theta_{i,c}}{x_i}\right)^{r_c}\right\}^{1/r_c},$$

$r_c \leq 0$, $\theta_{i,c} = 0$ if $i \notin c$, $\theta_{i,c} \geq 0$, $\sum_{c\in C}(-1)^{|c|}\theta_{i,c} \leq 1$ for each i.

Tilted Dirichlet (Coles and Tawn 1991)

A general construction: Suppose h^* is an arbitrary positive function on $\mathcal{S}_d$ with $m_d = \int_{\mathcal{S}_D} u_d h^*(\mathbf{u}) d\mathbf{u} < \infty$, then define

$$h(\mathbf{w}) = \left(\sum m_k w_k\right)^{-(D+1)} \prod_{d=1}^{D} m_d h^* \left(\frac{m_1 w_1}{\sum m_k w_k}, ..., \frac{m_D w_D}{\sum m_k w_k} \right).$$

h is density of positive measure H, $\int_{\mathcal{S}_D} u_d dH(\mathbf{u}) = 1$ each d.

As a special case of this, they considered Dirichlet density

$$h^*(\mathbf{u}) = \frac{\Gamma(\sum \alpha_j)}{\prod_d \Gamma(\alpha_d)} \prod_{d=1}^{D} u_d^{\alpha_d - 1}.$$

Leads to

$$h(\mathbf{w}) = \prod_{d=1}^{D} \frac{\alpha_d}{\Gamma(\alpha_d)} \cdot \frac{\Gamma(\sum \alpha_d + 1)}{(\sum \alpha_d w_d)^{D+1}} \prod_{d=1}^{D} \left(\frac{\alpha_d w_d}{\sum \alpha_k w_k} \right)^{\alpha_d - 1}.$$

Disadvantage: need for numerical integration

VII.1.b ESTIMATION

- Parametric
- Non/semi-parametric

Both approaches have problems, e.g. nonregular behavior of MLE even in finite-parameter problems; curse of dimensionality if D large.

Typically proceed by transforming margins to unit Fréchet first, though there are advantages in joint estimation of marginal distributions and dependence structure (Shi 1995)

VII.2 ALTERNATIVE FORMULATIONS OF MULTIVARIATE EXTREMES

- Ledford-Tawn-Ramos approach

- Heffernan-Tawn approach

The first paper to suggest that multivariate extreme value theory (as defined so far) might not be general enough was Ledford and Tawn (1996).

Suppose (Z_1, Z_2) are a bivariate random vector with unit Fréchet margins. Traditional cases lead to

$$\Pr\{Z_1 > r,\ Z_2 > r\} \sim \begin{cases} \text{const.} \times r^{-1} & \text{dependent cases} \\ \text{const.} \times r^{-2} & \text{exact independent} \end{cases}$$

The first case covers all bivariate extreme value distributions except the independent case. However, Ledford and Tawn showed by example that for a number of cases of practical interest,

$$\Pr\{Z_1 > r,\ Z_2 > r\} \sim \mathcal{L}(r) r^{-1/\eta},$$

where $\mathcal{L}$ is a slowly varying function ($\frac{\mathcal{L}(rx)}{\mathcal{L}(r)} \to 1$ as $r \to \infty$) and $\eta \in (0, 1]$.

Estimation: used fact that $1/\eta$ is Pareto index for $\min(Z_1, Z_2)$.

More general case (Ledford and Tawn 1997):

$$\Pr\{Z_1 > z_1, Z_2 > z_2,\} \;=\; \mathcal{L}(z_1, z_2) z_1^{-c_1} z_2^{-c_2},$$

$0 < \eta \le 1$; $c_1 + c_2 = \frac{1}{\eta}$; $\mathcal{L}$ slowly varying in sense that

$$g(z_1, z_2) \;=\; \lim_{t\to\infty} \frac{\mathcal{L}(tz_1, tz_2)}{\mathcal{L}(t,t)}.$$

They showed $g(z_1, z_2) = g_* \left(\frac{z_1}{z_1+z_2}\right)$ but were unable to estimate g_* directly — needed to make parametric assumptions about this.

More recently, Resnick and co-authors were able to make a more rigorous mathematical theory using concept of *hidden regular variation* (see e.g. Resnick 2002, Maulik and Resnick 2005, Heffernan and Resnick 2005; see also Section 9.4 of Resnick (2007)).

Ramos-Ledford (2009, 2008) approach:

$$\begin{aligned}
\Pr\{Z_1 > z_1, Z_2 > z_2,\} &= \tilde{\mathcal{L}}(z_1, z_2)(z_1 z_2)^{-1/(2\eta)}, \\
\lim_{u\to\infty} \frac{\tilde{\mathcal{L}}(ux_1, ux_2)}{\tilde{\mathcal{L}}(u, u)} &= g(x_1, x_2) \\
&= g_*\left(\frac{x_1}{x_1 + x_2}\right)
\end{aligned}$$

Limiting joint survivor function

$$\begin{aligned}
\Pr\{X_1 > x_1, X_2 > x_2\} &= \int_0^1 \eta \left\{\min\left(\frac{w}{x_1}, \frac{1-w}{x_2}\right)\right\}^{1/\eta} dH_\eta^*(w), \\
1 &= \eta \int_0^1 \{\min(w, 1-w)\}^{1/\eta} dH_\eta^*(w).
\end{aligned}$$

Multivariate generalization to $D > 2$:

$$\begin{aligned} \mathrm{Pr}\{X_1 > x_1, ..., X_D > x_D\} &= \int_{\mathcal{S}_D} \eta \left\{ \min_{1 \le d \le D} \left(\frac{w_d}{x_d} \right) \right\}^{1/\eta} dH_\eta^*(w), \\ 1 &= \int_{\mathcal{S}_D} \eta \left\{ \min_{1 \le d \le D} w_j \right\}^{1/\eta} dH_\eta^*(w). \end{aligned}$$

Open problems:

- How to find sufficiently rich classes of $H_\eta^*(w)$ to derive parametric families suitable for real data

- How to do the subsequent estimation

The Heffernan-Tawn approach

Reference:

A conditional approach for multivariate extreme values, by Janet Heffernan and Jonathan Tawn, *J.R. Statist. Soc. B* **66**, 497–546 (with discussion)

A set $C \in \mathcal{R}^d$ is an "extreme set" if

- we can partition $C = \cup_{i=1}^d C_i$ where

$$C_i = C \cap \left\{ x \in \mathcal{R}^d : \; F_{X_i}(x_i) > F_{X_j}(x_j) \right\} \text{ for all } i \neq j$$

  (C_i is the set on which X_i is "most extreme" among $X_1, ..., X_d$, extremeness being measured by marginal tail probabilities)

- The set C_i satisfies the following property: there exists a v_{X_i} such that

$$X \in C_i \text{ if and only if } X \in C_i \text{ and } X_i > v_{X_i}.$$

  (or: if $\mathbf{X} = (X_1, ..., X_d) \in C_i$ then so is any other $\mathbf{X}$ for which X_i is more extreme)

The objective of the paper is to propose a general methodology for estimating probabilities of extreme sets.

Marginal distributions

- Standard "GPD" model fit: define a threshold u_{X_i} and assume

$$\Pr\{X_i > u_{X_i} + x \mid X_i > u_{X_i}\} = \left(1 + \xi_i \frac{x}{\beta_i}\right)_+^{-1/\xi_i}$$

  for $x > 0$.

- Also estimate $F_{X_i}(x)$ by empirical CDF for $x < u_{X_i}$.

- Combine together for an estimate $\widehat{F}_{X_i}(x)$ across the entire range of x.

- Apply componentwise probability integral transform — $Y_i = -\log[-\log\{\widehat{F}_{X_i}(X_i)\}]$ to have approximately Gumbel margins (i.e. $\Pr\{Y_i \le y_i\} \approx \exp(-e^{-y})$).

- Henceforth assume marginal distributions are exactly Gumbel and concentrate on dependence among $Y_1, ..., Y_d$.

Existing techniques

Most existing extreme value methods with Gumbel margins reduce to

$$\Pr\{\mathbf{Y} \in t + A\} \approx e^{-t/\eta_{\mathbf{Y}}} \Pr\{\mathbf{Y} \in A\}$$

for some $\eta_{\mathbf{Y}} \in (0, 1]$. Ledford-Tawn classification:

- $\eta_{\mathbf{Y}} = 1$ is *asymptotic dependence* (includes all conventional multivariate extreme value distributions)
- $\frac{1}{d} < \eta_{\mathbf{Y}} < 1$ — positive extremal dependence
- $0 < \eta_{\mathbf{Y}} < \frac{1}{d}$ — negative extremal dependence
- $\eta_{\mathbf{Y}} = \frac{1}{d}$ — near extremal independence

Disadvantage: doesn't work for extreme sets that are not simultaneously extreme in all components

The key assumption of Heffernan-Tawn

Define $\mathbf{Y}_{-i}$ to be the vector $\mathbf{Y}$ with i'th component omitted.

We assume that for each y_i, there exist vectors of normalizing constants $\mathbf{a}_{|i}(y_i)$ and $\mathbf{b}_{|i}(y_i)$ and a limiting $(d-1)$-dimensional CDF $G_{|i}$ such that

$$\lim_{y_i \to \infty} \Pr\{\mathbf{Y}_{-i} \le \mathbf{a}_{|i}(y_i) + \mathbf{b}_{|i}(y_i)\mathbf{z}_{|i}\} \;=\; G_{|i}(\mathbf{z}_{|i}). \tag{4}$$

Put another way: as $u_i \to \infty$ the variables $Y_i - u_i$ and

$$\mathbf{Z}_{-i} \;=\; \frac{\mathbf{Y}_{-i} - \mathbf{a}_{|i}(Y_i)}{\mathbf{b}_{|i}(Y_i)}$$

are asymptotically independent, with distributions unit exponential and $G_{|i}(\mathbf{z}_{|i})$.

Examples in case $d = 2$

Distribution	η	$a(y)$	$b(y)$
Perfect pos. dependence	1	y	1
Bivariate EVD	1	y	1
Bivariate normal, $\rho > 0$	$\frac{1+\rho}{2}$	$\rho^2 y$	$y^{1/2}$
Inverted logistic, $\alpha \in (0, 1]$	$2^{-\alpha}$	0	$y^{1-\alpha}$
Independent	$\frac{1}{2}$	0	1
Morganstern	$\frac{1}{2}$	0	1
Bivariate normal, $\rho < 0$	$\frac{1+\rho}{2}$	$-\log(\rho^2 y)$	$y^{-1/2}$
Perfect neg. dependence	0	$-\log(y)$	1

Key observation: in all cases $b(y) = y^b$ for some b and $a(y)$ is either 0 or a linear function of y or $-\log y$ (for more precise conditions see equation (3.8) of the paper)

The results so far suggest the *conditional dependence model* (Section 4.1) where we assume the asymptotic relationships conditional on $Y_i = y_i$ are exact above a given threshold u_{Y_i}, or in other words

$$\Pr\{\mathbf{Y}_{-i} < \mathbf{a}_{|i}(y_i) + \mathbf{b}_{|i}(y_i)\mathbf{z}_{|i} \mid Y_i = y_i\} \;=\; G_{|i}(\mathbf{z}_{|i}), \quad y_i > u_{Y_i}.$$

Here $\mathbf{a}_{|i}(y_i)$ and $\mathbf{b}_{|i}(y_i)$ are assumed to be parametrically dependent on y_i through one of the alternative forms given by (3.8), and $G_{|i}(\mathbf{z}_{|i})$ is estimated nonparametrically from the empirical distribution of the standardized variables

$$\hat{\mathbf{Z}}_{-i} \;=\; \frac{\mathbf{Y}_{-i} - \hat{\mathbf{a}}_{|i}(y_i)}{\hat{\mathbf{b}}_{|i}(y_i)} \text{ for } Y_i = y_i > u_{Y_i}.$$

N.B. The threshold u_{Y_i} does not have to correspond to the threshold u_{X_i} used for estimating the marginal GPDs.

Some issues raised by this representation:

- Self-consistency of separate conditional models? (Section 4.2)

- Extrapolation — critical to have parametric forms for $\mathbf{a}_{|i}(y_i)$ and $\mathbf{b}_{|i}(y_i)$ (Section 4.3)

- Diagnostics — combine standard diagnostics for marginal extremes with tests of independence of $\mathbf{Z}_{-i}$ and Y_i. Also test whether the separate components of $\mathbf{Z}_{-i}$ are independent, since estimation via empirical distribution is much simpler in this case.

Inference

1. *Estimation of marginal parameters*

If ψ denotes the collection of all (β_i, ξ_i) parameters for the individual GPDs, maximize

$$\log L(\psi) \;=\; \sum_{i=1}^{d} \sum_{k=1}^{n_{u_{X_i}}} \log \widehat{f}_{X_i}(x_{i|i,k})$$

Here $n_{u_{X_i}}$ is the number of threshold exceedances in the i'th component and $\widehat{f}_{X_i}(x_{i|i,k})$ is the GP density evaluated at the k'th exceedance.

[In essence, if there is no functional dependence among the (β_i, ξ_i) for different i then this is just the usual marginal estimation of the GPD in each component. But if there is functional dependence, we estimate the parameters jointly by combining the individual likelihood estimation equations, ignoring dependence among the components.]

2. *Single conditional*

i.e. How do we estimate $\mathbf{a}_{|i}(y_i)$ and $\mathbf{b}_{|i}(y_i)$ for a single i, assuming parametric representation?

The problem: don't know the distribution of $\mathbf{Z}_{|i}$.

The solution: do it as if the $\mathbf{Z}_{|i}$ were Gaussian with known means $\boldsymbol{\mu}_{|i}$ and standard deviations $\boldsymbol{\sigma}_{|i}$

This leads to the formulae

$$\begin{aligned}\boldsymbol{\mu}_{|i}(y) &= \mathbf{a}_{|i}(y) + \boldsymbol{\mu}_{|i}\mathbf{b}_{|i}(y), \\ \boldsymbol{\sigma}_{|i}(y) &= \boldsymbol{\sigma}_{|i}\mathbf{b}_{|i}(y),\end{aligned}$$

and estimating equation

$$Q_i = -\sum_{j\neq i}\sum_{k=1}^{n_{u_{Y_i}}}\left[\log \sigma_{j|i}(y_{i|i,k}) + \frac{1}{2}\left\{\frac{y_{j|i,k} - \mu_{j|i}(y_{i|i,k})}{\sigma_{j|i}(y_{i|i,k})}\right\}^2\right].$$

3. *All conditionals*

To estimate all $\mathbf{a}_{|i}(y_i)$ and $\mathbf{b}_{|i}(y_i)$ jointly maximize

$$Q = \sum_{i=1}^{d} Q_i.$$

Analogous with *pseudolikelihood estimation* (Besag 1975)

4. *Uncertainty*

Bootstrap.....

Air quality application

The data: daily values of five air pollutants (O_3, NO_2, NO, SO_2, PM_{10}) in Leeds, U.K., during 1994–1998.

Two seasons: winter (NDJF), early summer (AMJJ)

Omit values around November 5 and some clear outliers

Marginal model: fit GPD above a (somewhat) high threshold, estimate 99% quantile with standard error (Table 4)

Table 4. Summary of generalized Pareto models fitted to the marginal distributions of the air pollution data†

Season	*Parameter*	*Results for the following pollutants:* O_3	NO_2	NO	SO_2	PM_{10}
Summer	u_{X_i}	43.0	66.1	43.0	22.0	46.0
	$\hat{F}_{X_i}(u_{X_i})$	0.9	0.7	0.7	0.85	0.7
	$\hat{\beta}_i$	15.8 (3.1)	9.1 (1.0)	32.2 (3.5)	42.9 (7.0)	22.8 (2.5)
	$\hat{\xi}_i$	−0.29 (0.14)	0.01 (0.08)	0.02 (0.07)	0.08 (0.12)	0.02 (0.08)
	$\hat{x}_i(0.99)$	70 (2)	75 (3)	180 (10)	152 (16)	127 (8)
Winter	u_{X_i}	28.0	151.6	49.0	23.0	53.0
	$\hat{F}_{X_i}(u_{X_i})$	0.7	0.7	0.7	0.7	0.7
	$\hat{\beta}_i$	6.2 (0.7)	9.3 (0.9)	117.4 (13.1)	19.7 (2.4)	37.5 (4.2)
	$\hat{\xi}_i$	−0.37 (0.06)	−0.03 (0.08)	−0.09 (0.08)	0.11 (0.09)	−0.20 (0.07)
	$\hat{x}_i(0.99)$	40 (1)	80 (3)	494 (30)	104 (10)	145 (6)

†Thresholds used for marginal modelling are denoted u_{X_i}; the associated non-exceedance probabilities are $\hat{F}_{X_i}(u_{X_i})$; estimated scale $\hat{\beta}_i$ and shape $\hat{\xi}_i$ parameters; estimated 0.99 quantiles $\hat{x}_i(0.99)$. Bootstrap-based standard errors are given in parentheses.

Heffernan and Tawn (2004)

Dependence model

Transform margins to Gumbel, select threshold for dependence modeling. Selected to be 70% quantile (for all five variables)

Estimate $(a_{j|i}, b_{j|i})$ and $(a_{i|j}, b_{i|j})$ for each combination of $i \neq j$, with sampling variability represented by convex hull of 100 bootstrap simulations (Fig. 5).

Several pairs do not exhibit weak pairwise exchangeability (e.g. PM_{10}, O_3 in summer)

Components of $\mathbf{Z}_{|i}$ are typically dependent

Some pairs exhibit negative dependence (e.g. O_3, NO in winter — consistent with chemical reactions)

Fig. 6 shows pseudosamples of other variables given NO over threshold ($C^5(23)$ are points for which sum of all 5 variables on Gumbel scale exceeds 23)

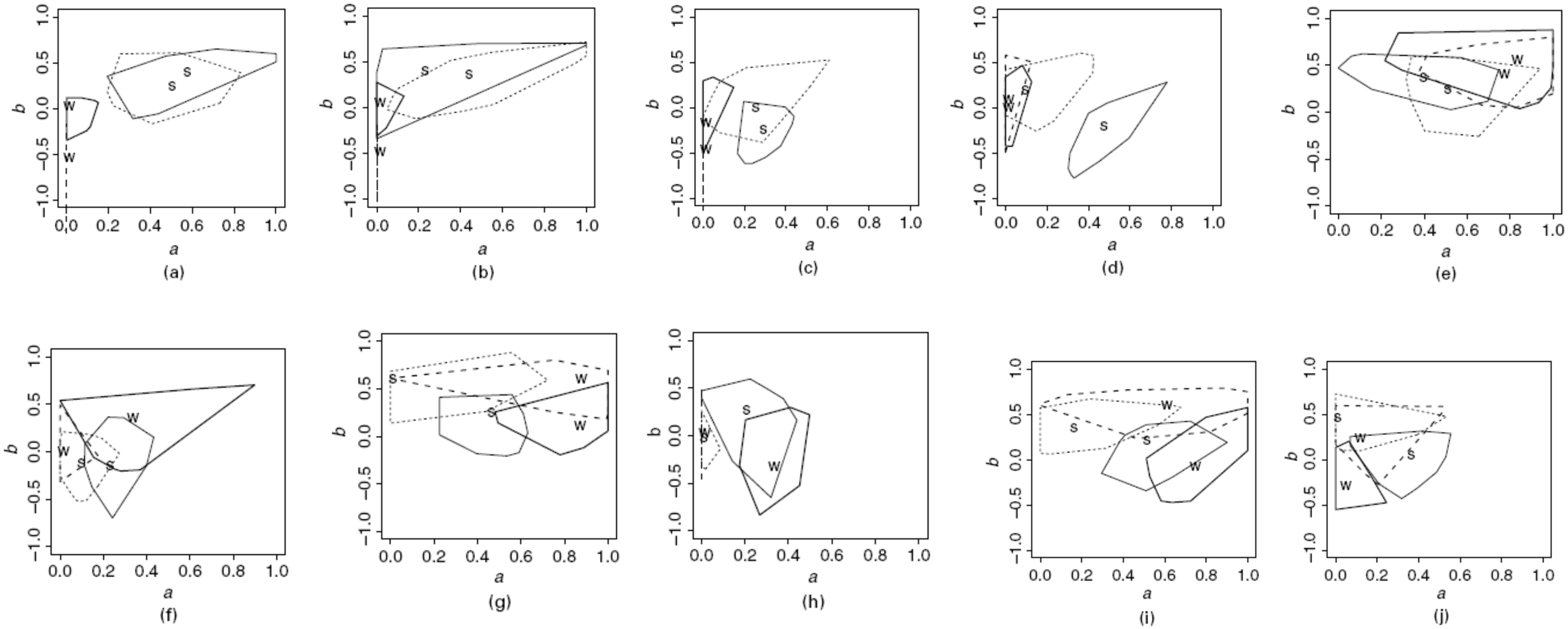


Fig. 5. Comparison of dependence parameter estimates $(\hat{a}_{i|j}, \hat{b}_{i|j})$ for (a) (NO_2, O_3), (b) (NO, O_3), (c) (SO_2, O_3), (d) (PM_{10}, O_3), (e) (NO, NO_2), (f) (SO_2, NO_2), (g) (PM_{10}, NO_2), (h) (SO_2, NO), (i) (PM_{10}, NO) and (j) (PM_{10}, SO_2), using a dependence threshold equal to the 70% marginal quantile: for i and j in the same order as the variables in the descriptor for each part of the figure, bootstrap convex hulls were used for $(\hat{a}_{i|j}, \hat{b}_{i|j})$ (———, summer; ━━━, winter) and for $(\hat{a}_{j|i}, \hat{b}_{j|i})$ (·······, summer; -------, winter) (associated point estimates: s, summer; w, winter)

Heffernan and Tawn (2004)

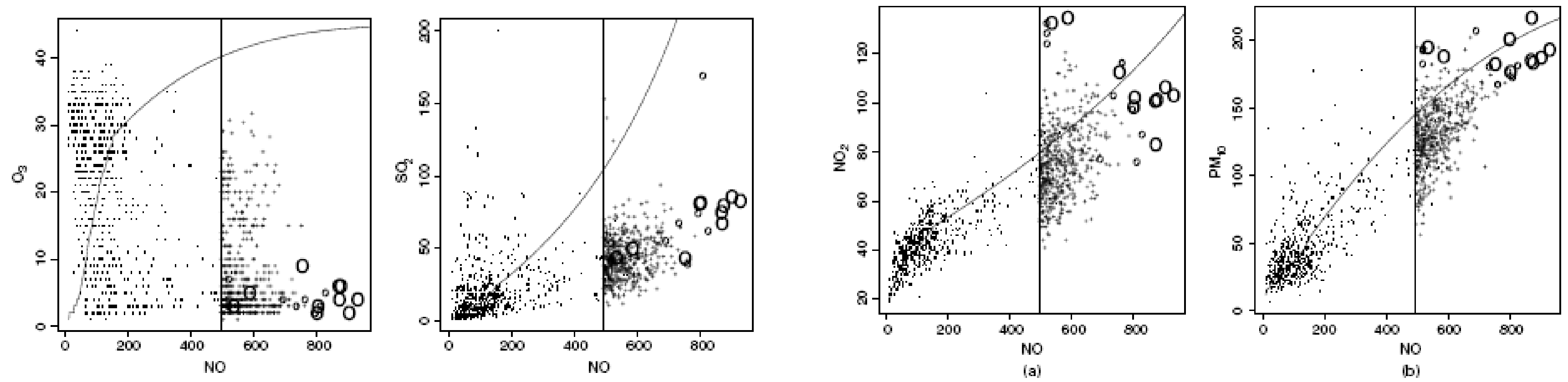


Fig. 6. Simulated winter air pollution points conditional on the NO component exceeding v_{X_i}, the 0.99 marginal quantile of this variable: |, threshold v_{X_i} (points below and above this threshold are the original data and data simulated under the fitted model respectively); +, points that do not fall in the set $C^5(23)$; $\circ$, points that fall in the set $C^5(23)$; $\bigcirc$, 10 points with the largest values of $\Sigma_{i=1}^{5} y_i$; ———, equal marginal quantiles

Heffernan and Tawn (2004)

Estimation of critical functionals

Contrast "joint probability" with "structure variable" approach (Coles and Tawn 1994)

(a) Estimate conditional mean of each other variable given that NO is about 95% or 99% threshold (Table 5)

(b) Sums of variables on Gumbel scale, e.g. for a subset $\mathcal{M} \subseteq \{1, ..., d\}$ with $|\mathcal{M}| = m$, define $C^m(v) = \{\mathbf{y} : \sum_{i \in \mathcal{M}} y_i > v\}$; define p-quantile v_p by property Pr$\{\mathbf{Y} \in C(v_p)\} = p$.

Compute return-level estimates for $C^m(v)$ with $\mathcal{M}$ corresponding to (O_3,NO_2) and (NO_2,SO_2,PM_{10}) (Fig. 7)

Table 5. Empirical and model-based estimates of conditional expectations of the air pollution variables given values of NO in excess of a range of quantiles of that variable†

X_j	*Season*	$E(X_j)$, *empirical*	$E\{X_j \mid X_i > x_i(0.95)\}$		$E\{X_j \mid X_i > x_i(0.99)\}$, *model based*
			Empirical	*Model based*	
O_3	Winter	20.0 (0.5)	8.8 (1.4)	10.3 (1.1)	8.3 (1.2)
	Summer	32.0 (0.4)	35.9 (3.0)	34.4 (2.4)	39.6 (4.3)
NO_2	Winter	44.2 (0.5)	67.2 (2.5)	65.1 (2.2)	75.4 (4.4)
	Summer	37.6 (0.5)	57.5 (2.6)	54.6 (2.4)	62.2 (4.3)
NO	Winter	135.5 (4.4)	454.0 (13.0)	431.5 (23.2)	569.9 (45.2)
	Summer	55.2 (1.5)	161.2 (7.2)	157.6 (8.2)	213.5 (17.5)
SO_2	Winter	21.0 (0.9)	38.4 (3.7)	35.6 (4.0)	44.6 (6.7)
	Summer	17.4 (1.2)	36.6 (11.3)	36.9 (5.4)	48.5 (11.8)
PM_{10}	Winter	48.4 (1.2)	105.8 (5.2)	105.0 (4.7)	132.3 (8.2)
	Summer	41.1 (1.0)	72.9 (5.2)	66.3 (4.5)	83.7 (7.9)

†Standard errors are given in parentheses. Variable X_i is NO throughout.

Heffernan and Tawn (2004)

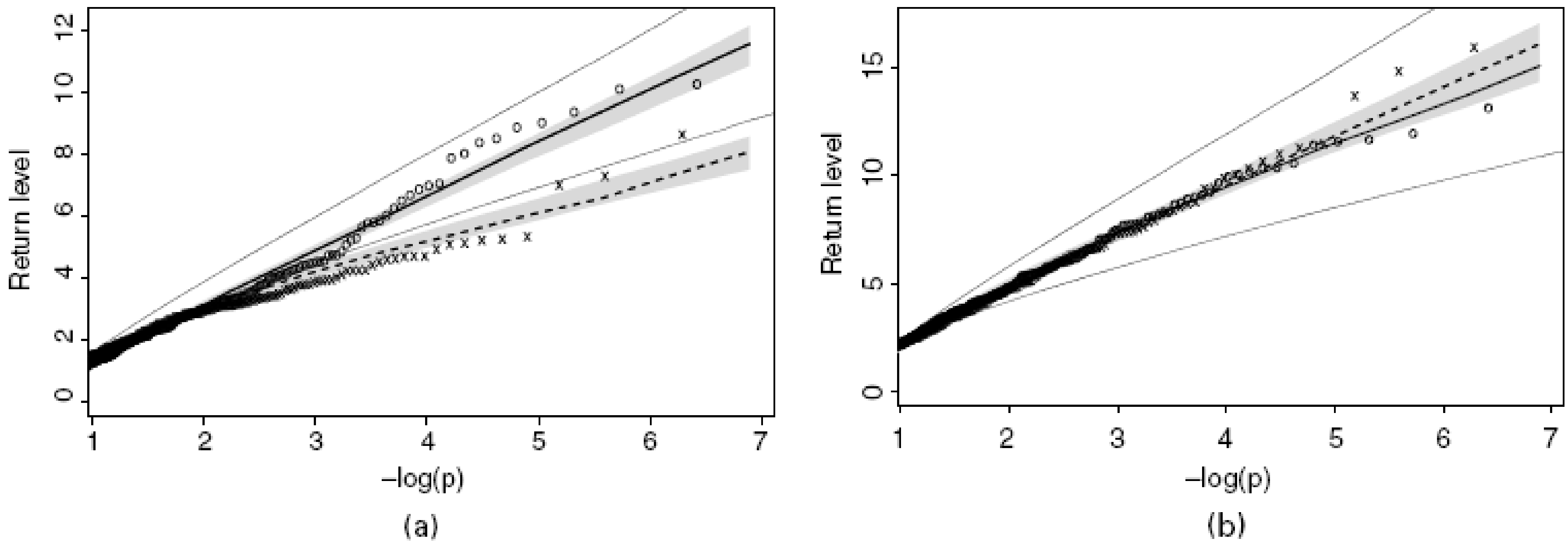


Fig. 7. Return level estimates for the set $C^m(v)$, for $\mathcal{M}$ corresponding to (a) (O_3, NO_2) and (b) (NO_2, SO_2, PM_{10}): ———, point estimates for summer; - - - - - - -, point estimates for winter variables; ■, pointwise 95% confidence intervals (which overlap in (b)); ∘, empirical points for summer; ×, empirical points for winter; ———, return levels calculated under perfect dependence (upper) and exact independence (lower)

Heffernan and Tawn (2004)

Further work on the Heffernan-Tawn model

See in particular

J.E. Heffernan and S.I. Resnick (2007), Limit laws for random vectors with an extreme component. *Annals of Applied Probability* **17**, 537–571

for a detailed analysis of the probability theory underlying the Heffernan-Tawn method. This should eventually lead to new statistical methods, though that has still to be done!

VII.3 MAX-STABLE PROCESSES

A stochastic process $\{Z_t,\ t \in T\}$, where T is an arbitrary index set, is called *max-stable* if there exist constants $A_{Nt} > 0$, B_{Nt} (for $N \geq 1,\ t \in T$) with the following property: if $Z_t^{(1)}, ..., Z_t^{(N)}$ are N independent copies of the process and

$$Z_t^* = \left(\max_{1 \leq n \leq N} Z_t^{(n)} - B_{Nt} \right) / A_{Nt}, \quad t \in T,$$

then $\{Z_t^*,\ t \in T\}$ is identical in law to $\{Z_t,\ t \in T\}$.

WLOG, assume margins are Unit Fréchet,

$$\Pr\{Z_t \leq z\} = e^{-1/z}, \quad \text{for all } t \tag{5}$$

in which case $A_{Nt} = N,\ B_{Nt} = 0$. Note that even though this assumption may be made for the purpose of characterizing the stochastic process, we would still have to estimate the marginal distributions in practice.

Construction of Max-Stable Processes

Let $\{(\xi_i, s_i),\ i \geq 1\}$ denote the points of a Poisson process on $(0, \infty) \times S$ with intensity measure $\xi^{-2} d\xi \times \nu(ds)$, where S is an arbitrary measurable set and ν a positive measure on S. Let $\{f(s,t),\ s \in S,\ t \in T\}$ denote a non-negative function for which

$$\int_S f(s,t)\nu(ds) = 1, \quad \text{for all } t \in T \tag{6}$$

and define

$$Z_t = \max_i \{\xi_i f(s_i, t)\}, \quad t \in T. \tag{7}$$

"Rainfall-storms" interpretation: S is space of "storm centres", ν as a measure representing distribution of storms over S. Each ξ_i represents the magnitude of a storm, and $\xi_i f(s_i, t)$ represents the amount of rainfall at position t from a storm of size ξ_i centred at s_i; the function f represents the "shape" of the storm. The max operation in (7) represents the notion that the observed maximum rainfall Z_t is a maximum over independent storms.

Fix $z_t > 0$ for each t, and consider the set

$$B = \{(\xi, s) : \quad \xi f(s,t) > z_t \text{ for at least one } t \in T\}.$$

The event $\{Z_t \le z_t \text{ for all } t\}$ occurs if and only if no points of the Poisson process lie in B. However, the Poisson measure of the set B is

$$\int_S \int_0^\infty I\left\{\xi > \min \frac{z_t}{(f(s,t)}\right\} \xi^{-2} d\xi \; \nu(ds) = \int_S \max_t \left\{\frac{f(s,t)}{z_t}\right\} \nu(ds)$$

where I is the indicator function, and consequently

$$\Pr\{Z_t \le z_t \;\text{ for all } t\} = \exp\left[-\int_S \max_t \left\{\frac{f(s,t)}{z_t}\right\} \nu(ds)\right]. \qquad (8)$$

It follows from (8) and (6) that the marginal distribution of Z_t, for any fixed t, is of Fréchet form (5), and by direct verification from (8), the process is max-stable.

Examples (Smith 1990)

Multivariate normal form:

Suppose $S = T = \Re^d$, ν is Lebesgue measure, and

$$f(s,t) = f_0(s-t) = (2\pi)^{-d/2}|\Sigma|^{-1} \exp\left\{-\frac{1}{2}(s-t)^T\Sigma^{-1}(s-t)\right\}.$$

Formula for $\Pr\{Y_{t_1} \le y_1, Y_{t_2} \le y_2\}$ is

$$\exp\left\{-\frac{1}{y_1}\Phi\left(\frac{a}{2}+\frac{1}{a}\log\frac{y_2}{y_1}\right)-\frac{1}{y_2}\Phi\left(\frac{a}{2}+\frac{1}{a}\log\frac{y_1}{y_2}\right)\right\}$$

where Φ is the standard normal distribution function and

$$a^2 = (t_1-t_2)^T\Sigma^{-1}(t_1-t_2).$$

Pickands dependence function:

$$A(w) = (1-w)\Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{1-w}{w}\right) + w\Phi\left(\frac{a}{2} + \frac{1}{a}\log\frac{w}{1-w}\right).$$

Limits $a \to 0$ and $a \to \infty$ become the extreme cases $A(w) = \max(w, 1-w)$ and $A(w) = 1$ representing, respectively, perfect dependence and independence.

Note that $A\left(\frac{1}{2}\right) = \Phi\left(\frac{a}{2}\right)$.

Remark. This dependence function was independently derived by Hüsler and Reiss (1989) as a limiting form for the joint distribution of bivariate extremes from a bivariate normal distribution with correlation ρ_n varying with sample size n, in such a way that $n \to \infty$, $\rho_n \to 1$ and $(1-\rho_n)\log n \to a^2/4$. They also gave a multivariate extension.

Multivariate t form:

$$f_0(x) = |\Sigma|^{-1/2}(\pi v)^{-d/2}\frac{\Gamma(v/2)}{\Gamma((v-d)/2)}\left(1+\frac{x^T\Sigma^{-1}x}{v}\right)^{-v/2},$$

valid for all $x \in \Re^d$, where Σ is again a positive definite covariance matrix and $v > d$.

No general formula for even the bivariate joint distributions, but we do have

$$A\left(\frac{1}{2}\right) = \frac{1}{2}\left\{1 + B\left(\frac{a^2}{a^2+4v^2};\frac{1}{2},\frac{v-d}{2}\right)\right\},$$

where $a = \sqrt{(t_1-t_2)^T\Sigma^{-1}(t_1-t_2)}$ again and B is the incomplete beta function,

$$B(y;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^y u^{\alpha-1}(1-u)^{\beta-1}du, \quad 0 \le y \le 1.$$

Note that, as $v \to \infty$, $A(1/2) \to \Phi(a/2)$ which is consistent with the MVN case.

Representations of Max-Stable Processes

The representation (8) first appeared in a number of papers in the 1980s, such as de Haan (1984), de Haan and Pickands (1986), Giné, Hahn and Vatan (1990).

However it is not the most general representation, as shown in particular by Schlather (2002).

Schlather's First Representation

Let Y be a measurable random function on $\Re^d$, $\mu = \mathsf{E}\left\{\int_{\Re^d} \max(0, Y(x))dx\right\} \in (0,\infty)$, Π a point process on $\Re^d \times (0,\infty)$ with intensity measure $d\Lambda(y,\xi) = \mu^{-1} dy \xi^{-2} d\xi$,

$$Z(x) = \sup_{(\xi,y)\in\Pi} \xi Y_{y,\xi}(x-y),$$

where each $Y_{y,\xi}(\cdot)$ is an independent copy of Y.

Then Z is max-stable with unit Fréchet margins.

Schlather's Second Representation

Let Y be a stationary process on $\Re^d$,
$\mu = \mathsf{E}\{\max(0, Y(x))\} \in (0, \infty)$, Π a point process on $(0, \infty)$ with intensity measure $\mu^{-1}\xi^{-2}d\xi$,

$$Z(x) \;=\; \max_{\xi \in \Pi} \xi Y_\xi(x) \;=\; \max_{\xi \in \Pi} \max\{0, \xi Y_\xi(x)\},$$

where each $Y_\xi(\cdot)$ is an independent copy of Y.

Then Z is max-stable with unit Fréchet margins.

Example:

Suppose Y is a stationary standard normal random field with correlation function $\rho(t)$. Let Π be a Poisson process on $(0, \infty)$ with intensity measure $d\Lambda(\xi) = \sqrt{2\pi}\xi^{-2}d\xi$. Then the process $Z(s) = \max_{\xi \in \Pi} \max\{0, \xi Y_\xi(s)\}$ defines an extremal Gaussian process. Its bivariate distributions satisfy

$$
\begin{aligned}
& \Pr\{Z(s_1) \le x_1, Z(s_2) \le x_2\} \\
= \;& \exp\left\{-\frac{1}{2}\left(\frac{1}{x_1} + \frac{1}{x_2}\right)\left(1 + \sqrt{1 - 2(\rho(s_1 - s_2) - 1)\frac{x_1 x_2}{(x_1 + x_2)^2}}\right)\right\},
\end{aligned}
$$

Another new class of bivariate EVDs!

Statistical Methods

Smith (1990)

Coles (1993)

Coles and Tawn (1996)

Buishand, de Haan and Zhou (2008)

Naveau, Guillou, Cooley and Diebolt (2009)

Method of Smith (1990)

Define the *extremal coefficient* between two variables: if X_1 and X_2 have same marginal cdf $F(\cdot)$,

$$\Pr\{X_1 \le x,\ X_2 \le x\} \ = \ F^{\theta}(x).$$

In BEVD case, $\theta = 2A\left(\frac{1}{2}\right)$.

Remark. Various people have proposed the extension where $F^{\theta}(x)$ is replaced by $F^{\theta(x)}(x)$, $\theta(x)$ depending on the level x. In some contexts, this could be equivalent to the Ledford-Tawn representation.

Smith proposed estimating the extremal coefficient between each pair of sites, $\tilde{\theta}_{ij}$ say, with standard error s_{ij}. Also, for the models considered so far there is a theoretical expression for θ_{ij} in terms of the parameters of the process. So we can use a nonlinear sum of squares criterion

$$\text{Minimize } S \;=\; \sum_{i,j} \left(\frac{\tilde{\theta}_{ij} - \hat{\theta}_{ij}}{s_{ij}} \right)^2$$

to determine the optimal $\hat{\theta}_{ij}$ as a parametric function of the max-stable process.

Coles (1993)

1. Rewrite model as

$$Z(t) \;=\; \max_i \{X_i f(S_i,t)\}\,,\; S_i \in \mathcal{S},\; t \in T,$$

   where (X_i, S_i) have Poisson measure $\mu(dx, ds) = x^{-2}dx\nu(ds)$ on $(0,\infty) \times \mathcal{S}$ (N.B. $\mathcal{S}$ and T don't have to coincide!)

2. Use all *storm profiles* rather than just pointwise maxima

3. Define a subset of stations $T_1 \subset T$ with $|T_1| = p$ and identify $\mathcal{S}$ with $\mathcal{S}_p$, the p-dimensional simplex

4. Let ν be the estimated H-measure on T_1 and $f(\mathbf{w}, t_j) = w_j$ for $\mathbf{w} \in \mathcal{S}_p,\; t_j \in T_1$

5. Extend $f(\mathbf{w}, t)$ to all $t \in T$.

Coles and Tawn (1996)

Objective: calculate extremal distributions of areal averages based on point rainfall data

Method:

1. Focus on i'th rain event with precipitation $X_i(v),\ v \in V$
2. Unit Fréchet Transformation $\Psi_v(X_i(v)) = \tilde{X}_i(v)$ after fitting $(mu(v), \psi(v), \xi(v))$ GEV model to each site v
3. Write $\tilde{X}_i(v) = \xi_i f(s_i, v)$ using point process representation of Coles (1993), fitted to storms for which at least one $X_i(v)$ exceeded some high threshold
4. Area average

$$Y_i = \frac{1}{\Delta_V} \int_V \Psi_v^{-1}(\xi_i f(s_i, v)) dv.$$

Naveau, Guillou, Cooley and Diebolt (2009)

Define *madogram*: if $M(x)$, $x \in X$ is a stationary process of pointwise maxima over some space X, with marginal cdf F, then

$$\nu(h) \;=\; \frac{1}{2}\mathsf{E}\left\{|F(M(x+h)) - F(M(x))|\right\}.$$

In particular, the extremal coefficient between two sites x and $x+h$ is $\frac{1+2\nu(h)}{1-2\nu(h)}$.

Disadvantage: only relevant to joint probabilities $\Pr\{M(x) \leq u,\; M(x+h) \leq v\}$ in cases $u = v$.

The λ-madogram:

$$\nu(h,\lambda) \;=\; \frac{1}{2}\mathsf{E}\left\{\left|F^{\lambda}(M(x+h)) - F^{1-\lambda}(M(x))\right|\right\}.$$

Then the exponent measure V is

$$V(\lambda, 1-\lambda) \;=\; \frac{c(\lambda)+\nu(h,\lambda)}{1-c(\lambda)-\nu(h,\lambda)},$$

where $c(\lambda) = 3/\{2(1+\lambda)(2-\lambda)\},\ \lambda \in (0,1)$.

If we assume the marginal cdf is known, then the estimator of $\nu(h,\lambda)$ is obvious. The main focus of the paper of Naveau *et al.* (2009) is to estimate $\nu(h,\lambda)$ when the marginal distribution is unknown. They do this nonparametrically using empirical process theory. This is an alternative to the Smith-Tawn-Coles method of estimating the distribution above a high threshold via GEV/GPD approximations.

Buishand, de Haan and Zhou (2008)

This will be Zhitao Zhang's class presentation. Therefore, I omit any discussion here, except to say that it is closely related to the methods of the previous papers.

Other (recent) books on extremes

1. Coles (2001)
   Best overall introduction to the whole subject
2. Finkenstadt and Rootzén (editors) (2003)
   Proceedings of European conference: includes RLS overview chapter, and Fougères on multivariate extremes
3. Beirlant, Goegebeur, Segers and Teugels (2004)
   Comprehensive overview of traditional extreme value theory, including multivariate extremes
4. De Haan and Ferreira (2006)
   Particular focus on max-stable ideas
5. Resnick (2007)
   Different focus: focus on probability models but including relevant statistics

REFERENCES

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004), *Statistics of Extremes: Theory and Appications* Wiley Series in Probability and Statistics, Chichester, U.K.

Besag, J.E. (1975), Statistical analysis of non-lattice data. *The Statistician* **24**, 179–195.

Buishand, T.A., de Haan, L. and Zhou, C. (2008), On spatial extremes: with application to a rainfall problem. *Annals of Applied Statistics* **2**, 624–642.

Coles, S.G. (1993), Regional modelling of extreme storms via max-stable processes. *JRSSB* **55**, 797–816.

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

Coles, S.G. and Tawn, J.A. (1991), Modelling extreme multivariate events. *J.R. Statist. Soc. B* **53**, 377–392.

Coles, S.G. and Tawn, J.A. (1994), Statistical Methods for Multivariate Extremes: An Application to Structural Design (with discussion). *Applied Statistics* **43**, 1–48.

Coles, S.G. and Tawn, J.A. (1996), Modelling extremes of the areal rainfall process. *JRSSB* **58**, 329–347.

Cooley, D., Nychka, D. and Naveau, P. (2007), Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association* **102**, 824–840.

Finkenstadt, B. and Rootzén, H. (editors) (2003), *Extreme Values in Finance, Telecommunications and the Environment.* Chapman and Hall/CRC Press, London.

Giné, E., Hahn, M. and Vatan, P. (1990), Max infinitely divisible and max stable sample continuous processes. *Probab. Theor. and Relat. Fields* **87**, 139–165.

Groisman, P.Ya, Knight, R.W., Easterling, D.A., Karl, T.R., Hegerl, G.C. and Razuvaev, V.N. (2005), Trends in intense precipitation in the climate record. *Journal of Climate* **18**, 1326–1350.

Gumbel, E.J. and Goldstein, N. (1964), Analysis of empirical bivariate extremal distributions. *J. Amer. Statist. Soc.* **59**, 794–816.

Haan, L. de (1984), A spectral representation for max-stable processes. *Ann. Probab.* **12**, 1194-1204.

Haan, L. de and Pickands, J. (1986), Stationary min-stable stochastic processes. *Prob. Th. Rel. Fields* **72**, 477-492.

Hüsler, J. and Reiss, R.-D. (1989), Maxima of normal random vectors: Between independence and complete dependence. *Statistics and Probability Letters* **7**, 283-286.

Haan, L. de and Ferreira, A. (2006), *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering, New York.

Heffernan, J.E. and Resnick, S. (2005), Hidden regular variation and the rank transform. *Advances in Applied Probabilitty* **37-2**, 393–414.

Heffernan, J.E. and Resnick, S. (2007), Limit laws for random vectors with an extreme component. *Annals of Applied Probability* **17**, 537–571

Heffernan, J.E. and Tawn, J.A. (2004), A conditional approach for multivariate extreme values. *J.R. Statist. Soc. B* **66**, 497–546 (with discussion)

Joe, H. (1990), Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statist. and Probab. Letters* **9**, 75–81.

Ledford, A.W. and Tawn, J.A. (1996), Modelling dependence within joint tail regions, *J.R. Statist. Soc. B* **59**, 475–499

Ledford, A.W. and Tawn, J.A. (1997), Statistics for near independence in multivariate extreme values. *Biometrika* **83**, 169–187.

Mannshardt-Shamseldin, E.C., Smith, R.L., Sain, S.R., Mearns, L.O. and Cooley, D. (2008), Downscaling extremes: A comparison of extreme value distributions in point-source and gridded precipitation data. Preprint, under revision.

Maulik, K. and Resnick, S. (2005), Characterizations and examples of hidden regular variation. *Extremes* **7**, 31–67

Naveau, P., Guillou, A., Cooley, D. and Diebolt, J. (2009), Modelling pairwise dependence of maxima in space. *Biometrika* **96**, 1–17.

Ramos, A. and Ledford, A. (2009), A new class of models for bivariate joint tails. *J.R. Statist. Soc. B* **71**, 219–241.

Ramos, A. and Ledford, A. (2008), An alternative point process framework for modeling multivariate extreme values. Preprint.

Resnick, S. (1987), *Extreme Values, Point Processes and Regular Variation.* Springer Verlag, New York.

Resnick, S. (2002), Second order regular variation and asymptotic independence. *Extremes* **5**, 303–336

Resnick, S. (2007), *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York.

Robinson, M.E. and Tawn, J.A. (1995), Statistics for exceptional athletics records. *Applied Statistics* **44**, 499–511.

Sang, H. and Gelfand, A.E. (2008), Hierarchical modeling for extreme values observed over space and time. Published online by *Environmental and Ecological Statistics*.

Schlather, M. (2002), Models for Stationary Max-Stable Random Fields. *Extremes* **5**(1), 33–44.

Shi, D. (1995), Fisher information for a multivariate extreme value distribution. *Biometrika* **82**, 644–649.

Smith, R.L. (1989), Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone (with discussion). *Statistical Science* **4**, 367–393.

Smith, R.L. (1990), Max-stable processes and spatial extremes. Unpublished, available from
http://www.stat.unc.edu/postscript/rs/spatex.pdf

Smith, R.L. (1997), Statistics for exceptional athletics records. Letter to the Editor, *Applied Statistics* **46**, 123–127.

Smith, R.L. (2003), Statistics of extremes, with applications in environment, insurance and finance. Chapter 1 of *Extreme Values in Finance, Telecommunications and the Environment*, edited by B. Finkenstädt and H. Rootzén, Chapman and Hall/CRC Press, London, pp. 1–78.

Smith, R.L. and Goodman, D. (2000), Bayesian risk analysis. Chapter 17 of *Extremes and Integrated Risk Management*, edited by P. Embrechts. Risk Books, London, 235-251.

Smith, R.L. and Weissman, I. (1994), Estimating the extremal index. *J.R. Statist. Soc. B* **56**, 515–128.

Stott, P.A., Stone, D.A. and Allen, M.R. (2004), Human contribution to the European heatwave of 2003. *Nature* **432**, 610–614 (December 2 2004)

Tawn, J.A. (1990), Modelling multivariate extreme value distributions. *Biometrika* **77**, 245–253.