Introduction to Extreme Value Analysis

Alexis Hannart
Ouranos and McGill University

29 October 2019

Based on lectures and supporting material provided by:

- Dan Cooley, Colorado State University, USA
- Anthony Davison, EPFL, Switzerland
- Philippe Naveau, CNRS, France
- and more.

Books on Extremes

- Coles (2001), *An Introduction to Statistical Modeling of Extreme Values*
- Beirlant et al. (2004) *Statistics of Extremes: Theory and Applications*
- de Haan and Ferreira (2006) *Extreme Value Theory*
- Resnick (2007) *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*

Software

R packages: `ismev`, `evd`, `evir`, `SpatialExtremes`, `extRemes`

Short Course Materials

`http://www.stat.colostate.edu/~cooleyd/MontrealJSM2013/`

Why study extremes?

Although infrequent, extremes have large human impact.

Colorado precipitation examples:

Big Thompson Flood, 1976

- 145 killed
- $41m damage

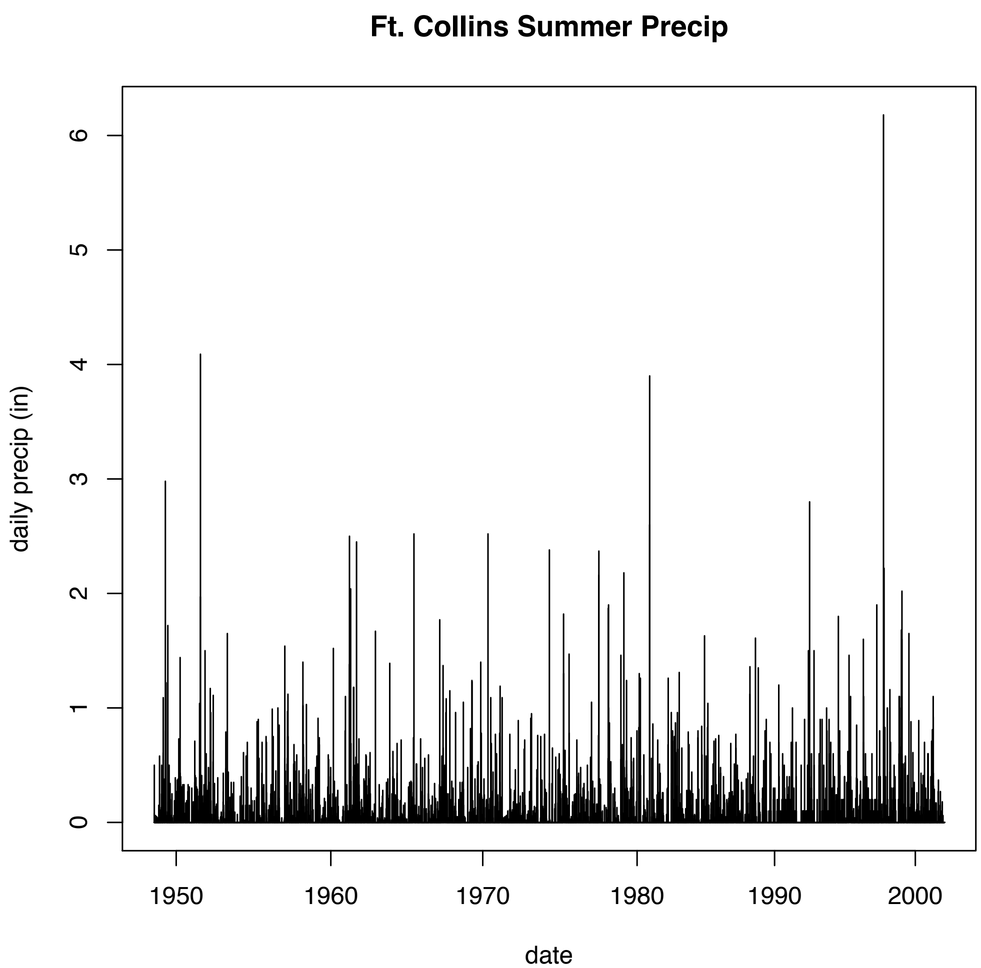
Eve Gruntfest, UCCS

Ft Collins Flood, 1997

- 5 killed
- $250m damage

John Weaver

Fort Collins Precipitation

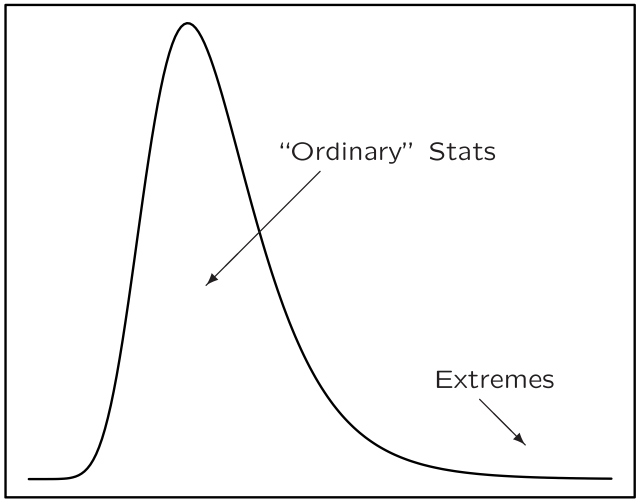


- Spike corresponds to 1997 event, recording station **not** at center of storm.
- Associated question: How unusual was event?

"Ordinary" vs Extreme Value Statistics

"Ordinary" Statistics: Describes main part of distribution.

Extremes: Characterizes the tail of the distribution.

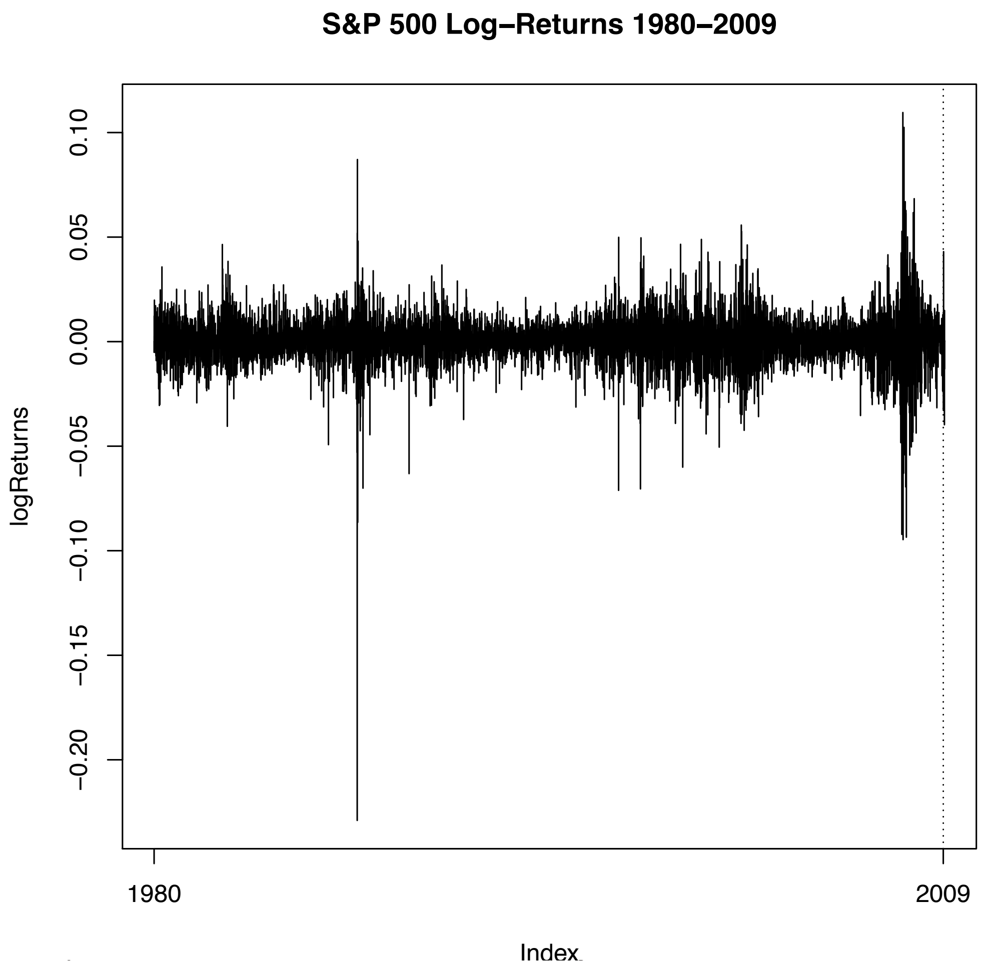


Why study extremes?

Application areas:

- hydrology (stream/river flows, flooding)
- climate variables: precipitation, wind, heat-waves, ...
- finance
- insurance/reinsurance
- engineering (structural design, failure)
- not much done (yet) in medicine, biology, ecology

S&P 500 Log-Returns



- Black Monday: Oct 19 1987; Volatility in 2008

Outline

Part I: Univariate block-maximum approaches.

Part II: Univariate threshold-exceedance approaches.

Part III: Other topics in univariate extremes analyses.

Part IV: Introduction to bivariate extremes.

Outline

Part I: Univariate block-maximum approaches.

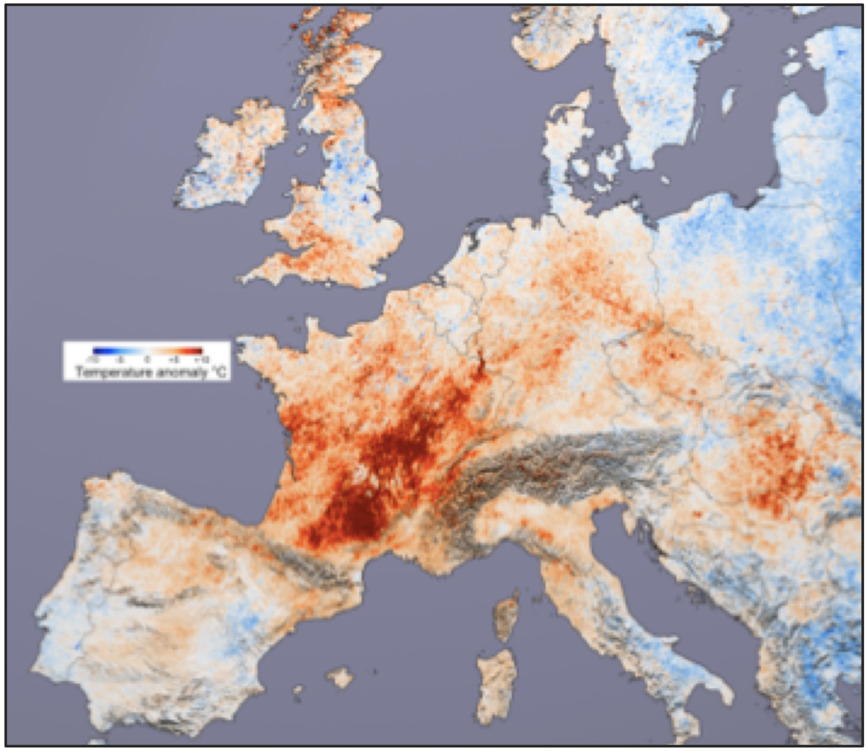
Part II: Univariate threshold-exceedance approaches.

Part III: Other topics in univariate extremes analyses.

Part IV: Introduction to bivariate extremes.

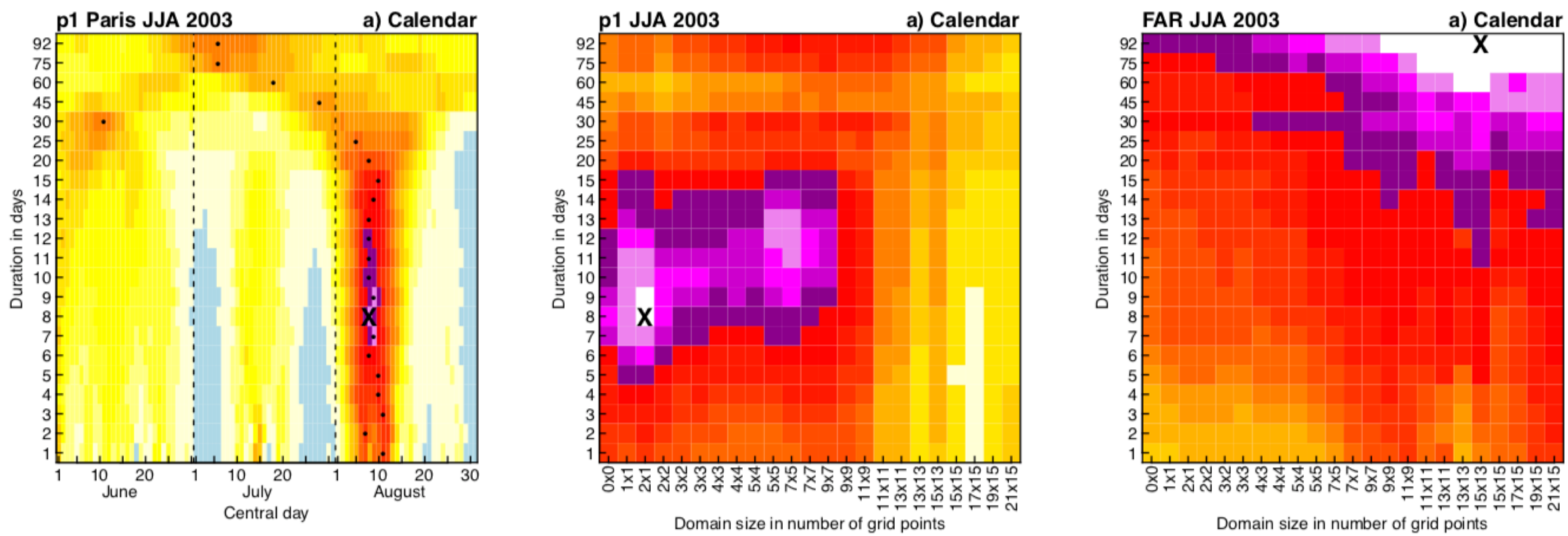
European 2003 heatwave: what was the event ?

August 2003 Tmax anomaly



- There are multiple ways to define an event, even for a heatwave
  - which variable, which threshold ?
  - which spatial and temporal scales ?

European 2003 heatwave: what was the event ?



p1 maximization wrt duration in Paris

p1 maximization wrt domain size around Paris

p1/p0 maximization wrt domain size around Paris

p1/p0 ~ 1000

From Ribes and Cattiaux

Outline

Part I: Univariate block-maximum approaches.

Part II: Univariate threshold-exceedance approaches.

Part III: Other topics in univariate extremes analyses.

$\boxed{\text{Part IV: Introduction to bivariate extremes.}}$

Part I: Univariate block-maximum approaches

- Illustrative case study: Fort Collins precipitation.
- Max-stability and the GEV distribution.
- Inference.
- Uncertainty estimation.
- Example.

Return Levels and Return Periods

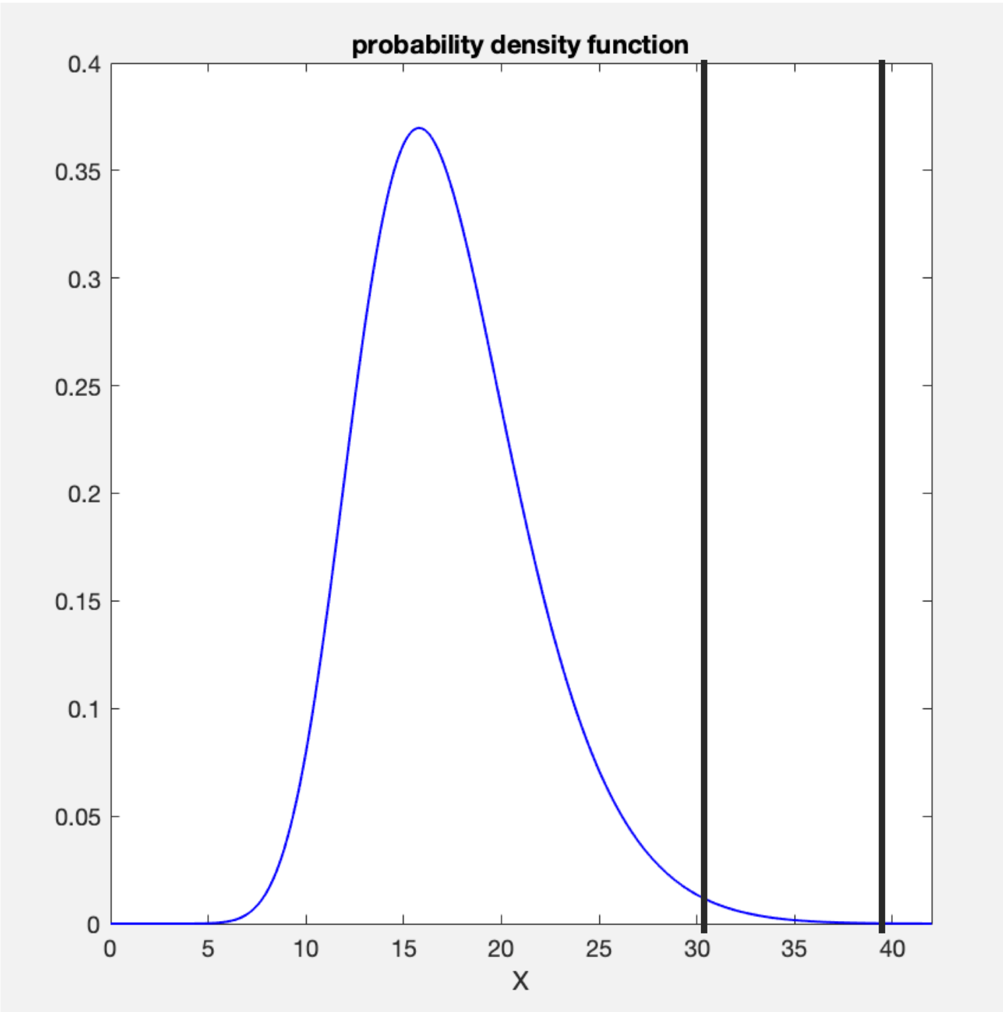
The m-year *return level* as the high quantile for which the probability that the *annual maximum* exceeds this quantile is $1/m$.

- Canonical example: “100-year flood”.
- Financial equivalent: ”value-at-risk”.

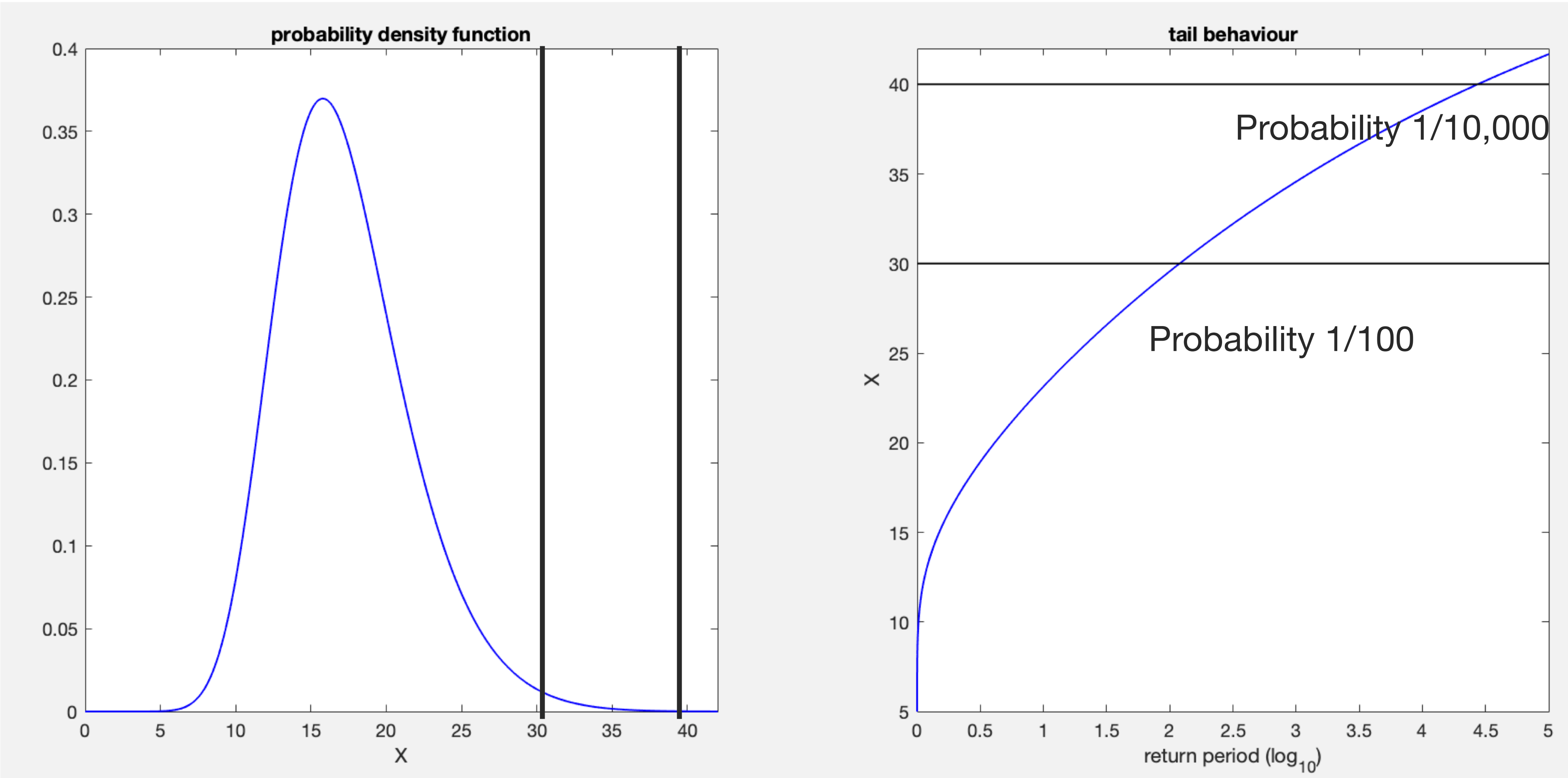
The *return period* of a particular event is the inverse of the probability that the event will be exceeded in any given year.

Both of these definitions assume stationarity. See Cooley (2012) and references therein for non-stationarity setting.

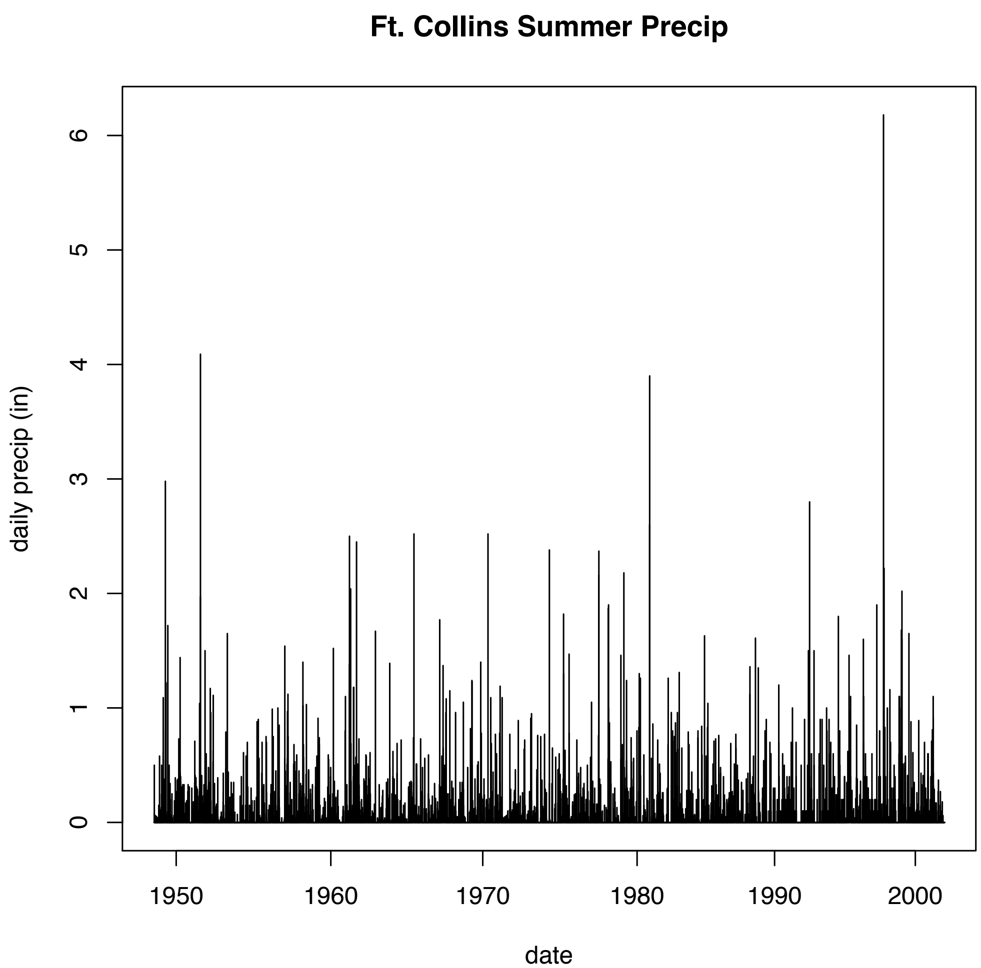
Probability Density Function (PDF)



PDF and return level plot



Fort Collins Precipitation



- Spike corresponds to 1997 event, recording station **not** at center of storm.
- Associated question: How unusual was event?

How unusual was the Fort Collins event?

Measured value for 1997 event: 6.18 inches.

Let's analyze data preceding the event (1948-1990) and estimate the 'return period' of an event of 6.18 inches.

We need to answer the question: "What is the probability the annual maximum event is larger than 6.18 inches?"

Also, estimate the '100-year return level'.

Both questions require extrapolation into the tail. Largest observation (1948-1990) is 4.09 inches.

Model the data in two ways:

1. Model *all* (non-zero) data.
2. Model only extreme data.

Note: `R` code for both analyses in `1BlockMaxima.R` script.

Estimating the probability of a rare event

- The easiest way: empirical frequency

$$\widehat{p} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\{x_i > u\}}$$

Estimating the probability of a rare event

- The easiest way: empirical frequency

$$\widehat{p} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{\{x_i > u\}}$$

- A very noisy estimate:

$$RE = \frac{1}{\sqrt{\widehat{p}\,N}}$$

To estimate a return time of 100 years with a 10% error, one needs 10,000 years of data.

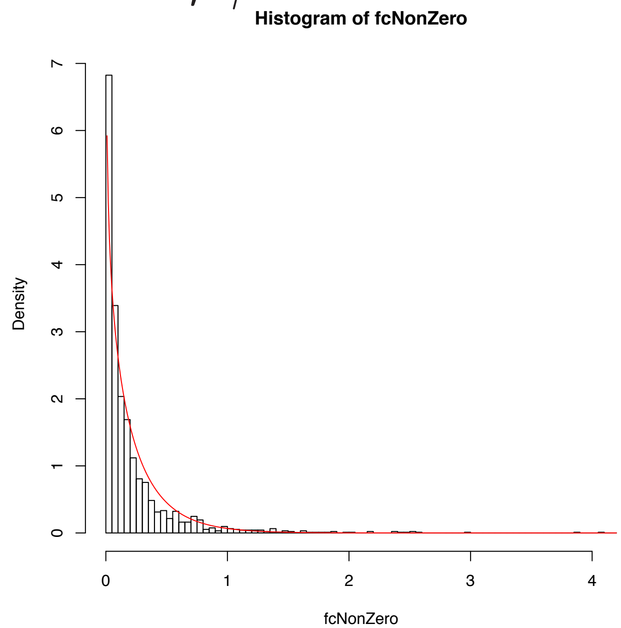
Modeling all precipitation data

Let X_t be the daily "summer" precipitation amount for Fort Collins. (Summer = Apr-Oct)

To model precipitation, we need to account for zeroes.

Assume: $\begin{cases} X_t > 0 & \text{w.p. } p \\ X_t = 0 & \text{w.p. } 1-p. \end{cases}$ $\quad \widehat{p} = 0.218.$

Further, assume that $[X_t \mid X_t > 0] \sim \text{Gamma}(\alpha, \beta)$.
ML estimates: $\widehat{\alpha} = 0.784$, $\widehat{\beta} = 3.52$.



All precipitation model estimate

$$\begin{aligned} P(X_t > 6.18) &= P(X_t > 6.18 \mid X_t > 0)P(X_t > 0) \\ &= (1 - F_X(6.18))(0.218) \\ &= 1.47 * 10^{-10}(0.218) = 3.20 * 10^{-11} \end{aligned}$$

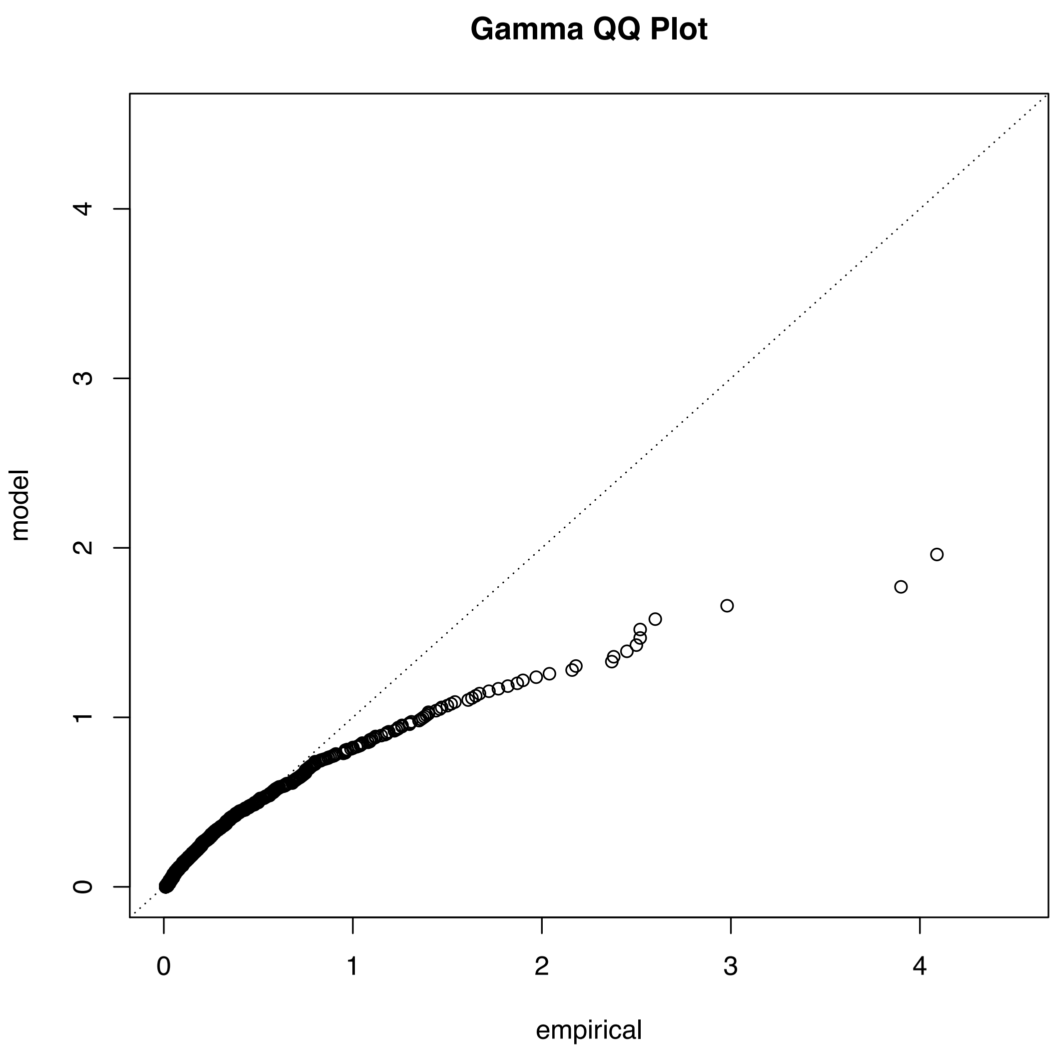
$$\begin{aligned} P(\text{ann max} > 6.18) &= 1 - P(\text{entire year's obs} < 6.18) \\ &= 1 - (1 - P(\text{indiv obs} > 6.18))^{214} \\ &= 1 - (1 - 3.20 * 10^{-11})^{214} \\ &= 6.86 * 10^{-9} \end{aligned}$$

(Assumes independence of daily observations, 214 “summer” days in a year.)

Return period $= (6.86 * 10^{-9})^{-1} = 145,815,245$ years.

Estimate of 100-year return level: 3.06 inches.

All precipitation model



Note: 98% of model's mass and 97% of data are < 1.

[redacted] the kernel estimator

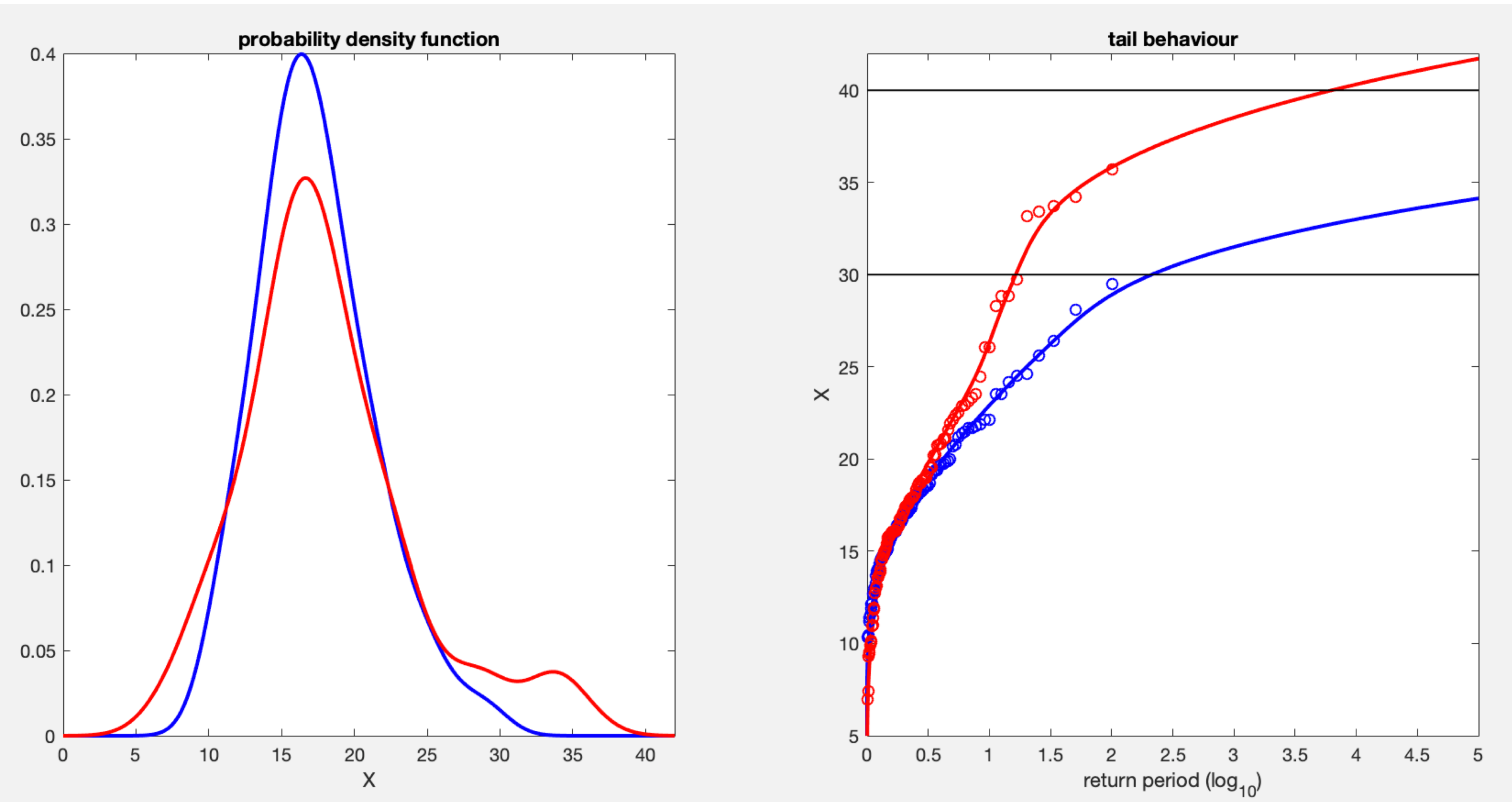
$$\widehat{p} = \frac{1}{N} \sum_{i=1}^{N} K\left(\frac{x_i - u}{h}\right)$$

- K is the so-called kernel:

$$K(x) = \int_x^{\infty} \mathcal{N}(z) \mathrm{d}z$$

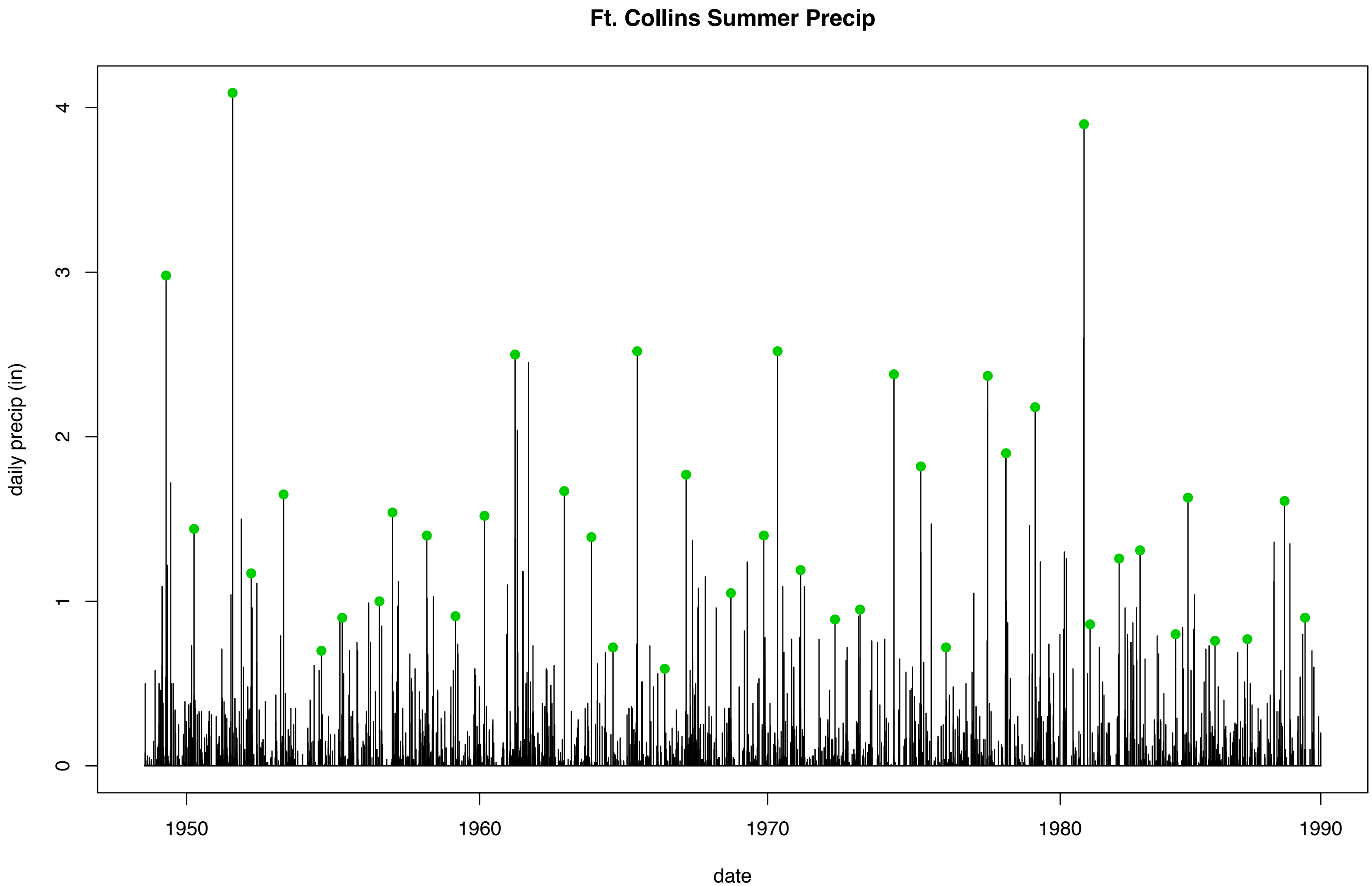
- It basically smoothes the estimation.

[redacted] N = 100



Modeling annual maxima

Approach: Extract annual maxima and fit a model.



Modeling annual maxima

Let $M_n = \max_{t=1,\ldots,n}(X_t)$. Assume $M_n \sim \text{GEV}(\mu, \sigma, \xi)$.
(We will discuss why the GEV is the right distribution later.)

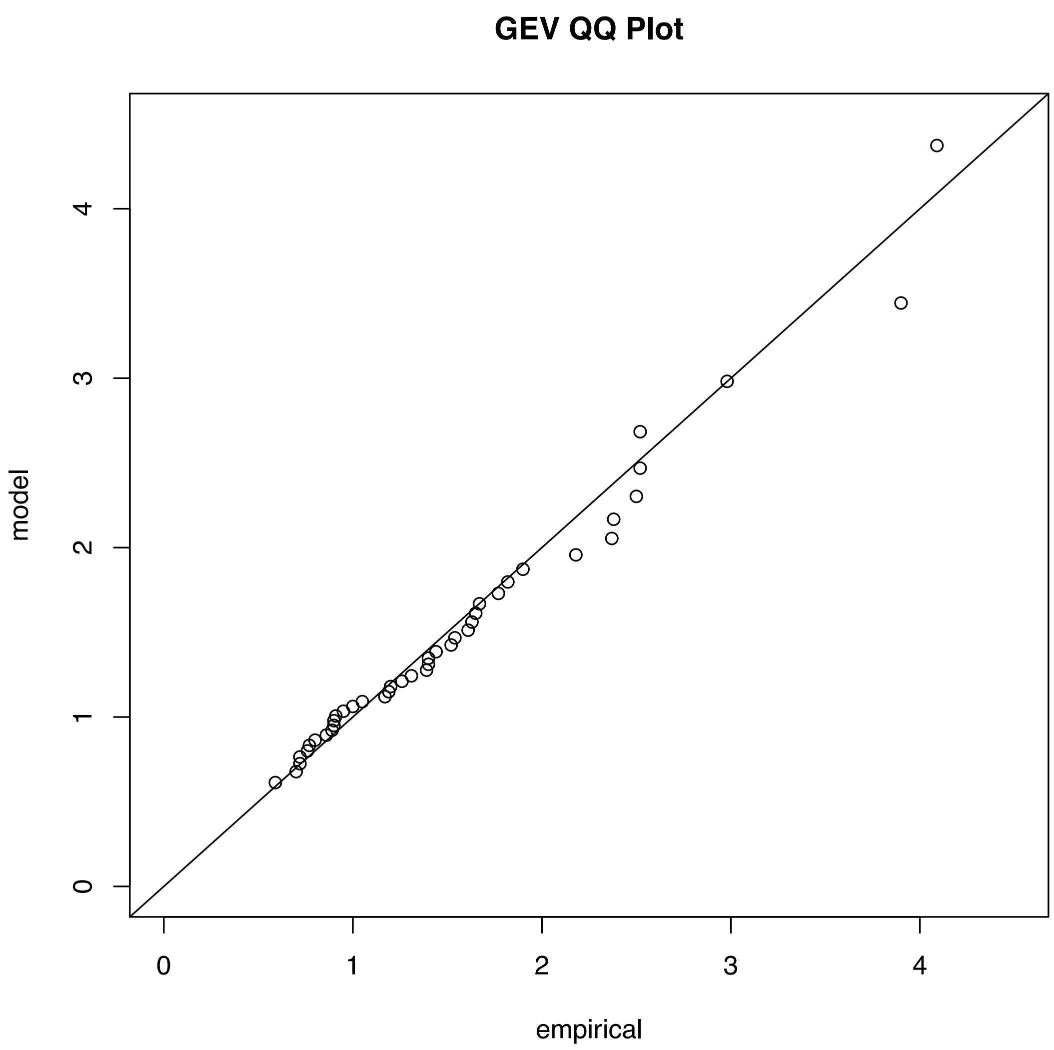
$$F_{M_n}(x) = P(M_n \leq x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}.$$

ML estimates: $\widehat{\mu} = 1.11$, $\widehat{\sigma} = 0.46$, $\widehat{\xi} = 0.31$.

$P(\text{ann max} > 6.18) = 1 - F_{M_n}(6.18) = 0.008$.
Return period point estimate: $0.008^{-1} = 121$ years.

100-year return level point estimate: 5.80 inches.

Modeling annual maxima



Note: Plot shows only annual maxima.

Why use only 'extreme' observations?

Two approaches for extracting extreme observations:

1. Block-maximum approach (done above)
2. Threshold-exceedance approach

Heuristic explanation: Phenomena which generate extreme observations are fundamentally different than those which generate typical observations.

Mathematical explanation: Assume X_t has cdf $F_X(x)$.

$$\begin{aligned} F_{M_n}(x) = P(M_n \leq x) &= P(X_t \leq x \text{ for all } t = 1, \ldots, n) \\ &= P^n(X_t \leq x) \\ &= F_X^n(x) \end{aligned}$$

If we know F_X exactly, then we know F_{M_n} exactly. But if we have to estimate F_X, any errors get amplified by n.

"Let the tails speak for themselves."

Why is the GEV the right distribution?

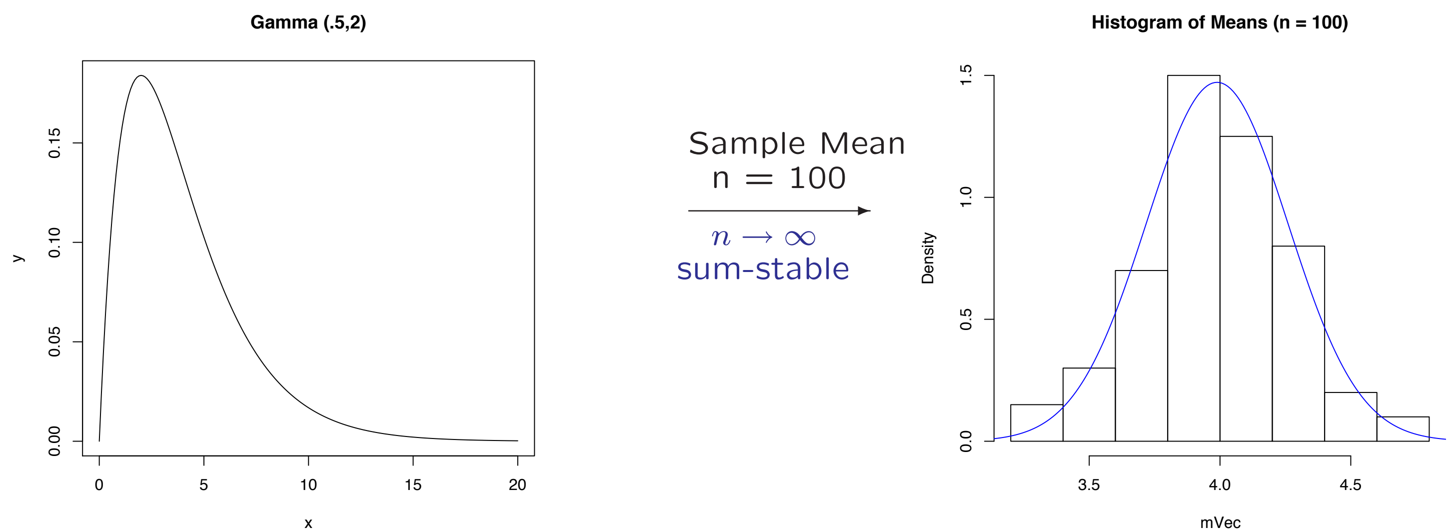
Answer: In a minute.

Why is the GEV the right distribution?

Answer: In a minute.

Why is the normal the right distribution for modeling sample means?

Answer: The central limit theorem.
The normal is (sum-)stable.



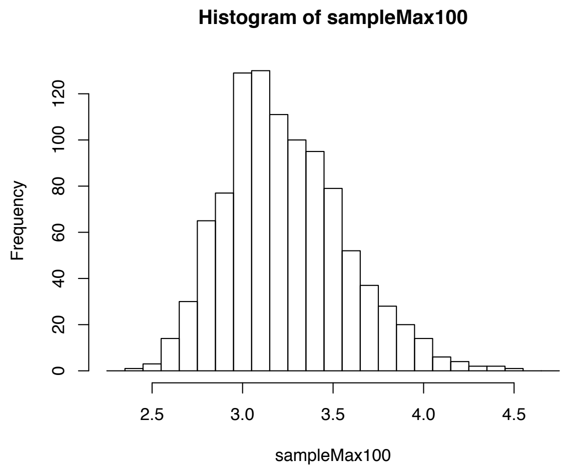
Important: We don't need information about the distribution of X_t to know about the distribution of the sample mean.

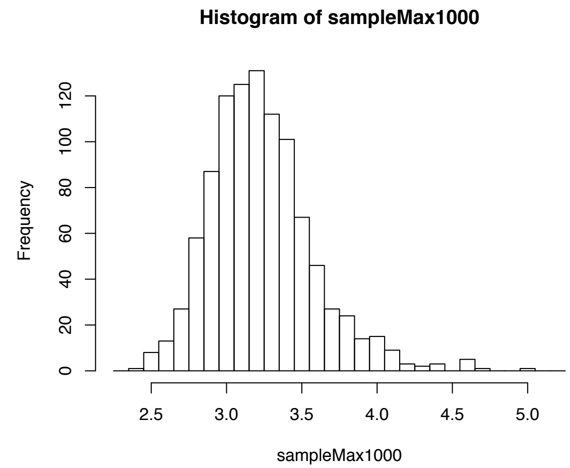
Why is the GEV the right distribution?

R Demo

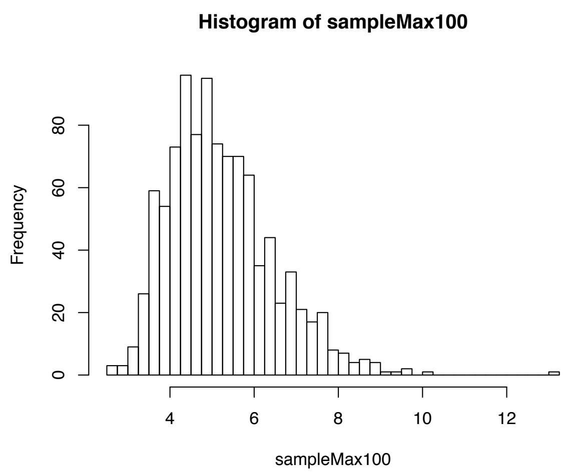
Distributions of sample maxima

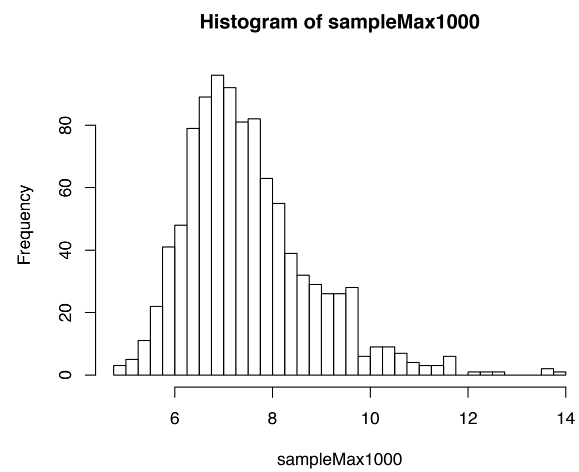
Normal





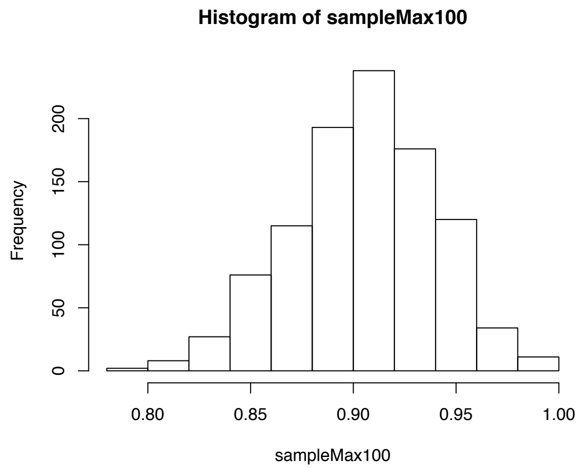
Exponential

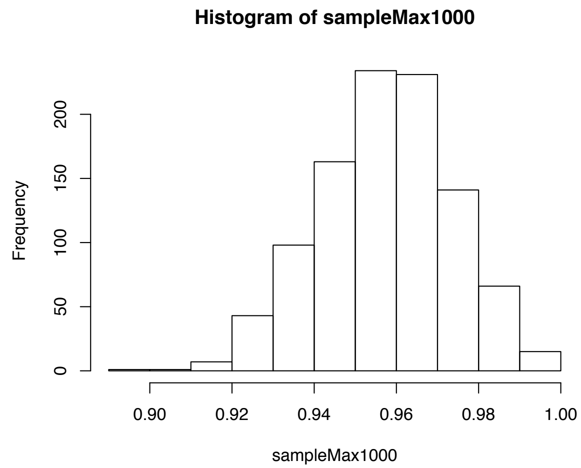




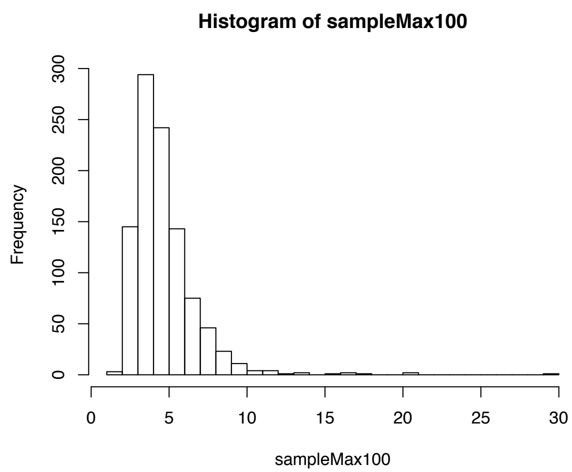
Distributions of sample maxima

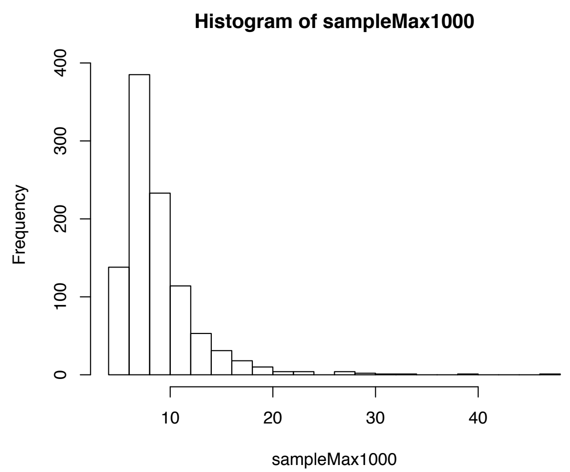
Beta





Student’s t (4 df)





Three-types Theorem

Like a Central-Limit Theorem for maxima.

Let $M_n = \max_{t=1,\ldots,n} X_t$, where X_t are iid. If there exist normalizing sequences a_n and b_n such that $P\left(\frac{M_n - b_n}{a_n} \leq x\right) \to G(x)$ (nondegenerate) as $n \to \infty$, then

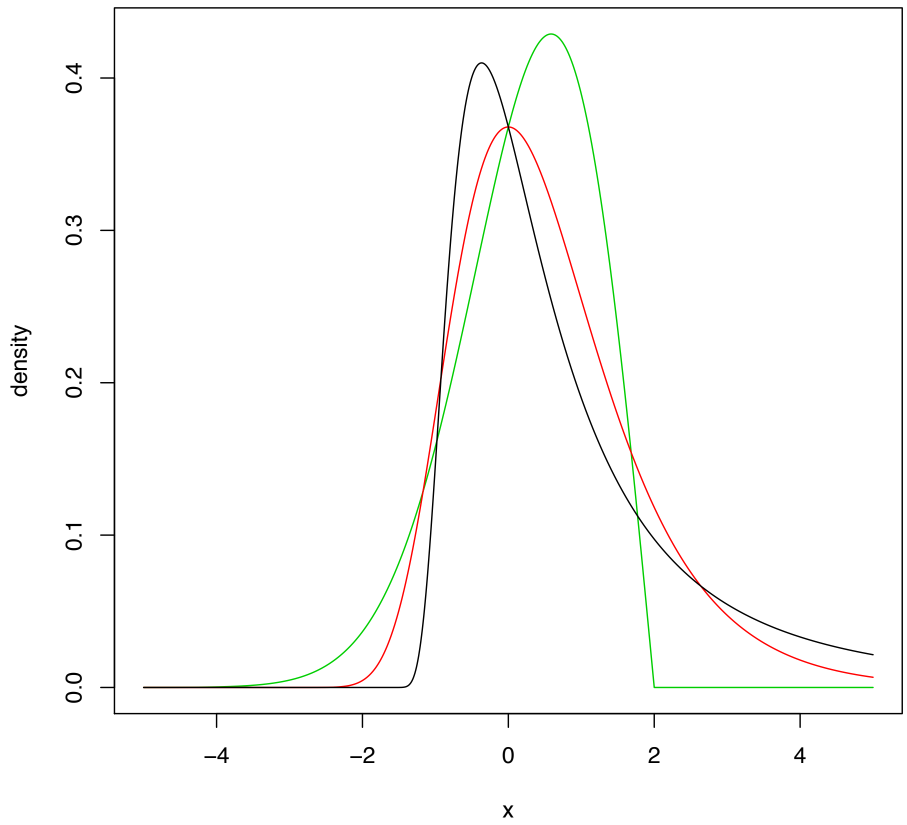
$$G(x) = \exp\left\{-\left[1 + \xi x\right]^{-1/\xi}\right\}.$$

ξ determines the tail behavior.

- $\xi \leq 0$: Weibull (or reverse Fréchet) case (bounded tail)
- $\xi = 0$: Gumbel case (light tail), interpreted as limit
- $\xi > 0$: Fréchet case (heavy tail)

Important: We don't need information about the distribution of X_t to know about the distribution of M_n.

Limiting Distributions



Weibull ($\xi = -0.5$)

Gumbel

Fréchet ($\xi = 0.5$)

Statistical Practice

Assume n is fixed and large enough so that:

$$
\begin{aligned}
P\left(\frac{M_n - b_n}{a_n} \leq x\right) &\approx \exp\left\{-\left[1+\xi x\right]^{-1/\xi}\right\} \\
\Rightarrow P\left(M_n \leq y\right) &\approx \exp\left\{-\left[1+\xi\left(\frac{y-b_n}{a_n}\right)\right]^{-1/\xi}\right\} \\
&= \exp\left\{-\left[1+\xi\left(\frac{y-\mu}{\sigma}\right)\right]^{-1/\xi}\right\},
\end{aligned}
$$

where y s.t. $1+\xi\left(\frac{y-\mu}{\sigma}\right) \geq 0$.

For the Fort Collins precipitation data, $n = 214$, the number of 'summer' days in a year.

Then, we have a three-parameter estimation problem. μ, σ, and ξ can be estimated via (numerical) maximum likelihood, or some method-of-moments type procedures.

Fort Collins Data Revisited

Parameters estimated by numerical ML (standard errors):

$$\widehat{\mu} = 1.11 \quad (0.086)$$
$$\widehat{\sigma} = 0.46 \quad (0.074)$$
$$\widehat{\xi} = 0.31 \quad (0.181)$$

- Estimate for ξ is (quite) heavy-tailed.
- Notice the large standard error on ξ.

Point estimate for the 100-year return level: 5.8 inches

95% confidence interval for 100-year return level

via delta-method: (1.2, 10.4)

via profile-likelihood: (3.5, 18.8)

GEV and return levels

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

Computing the return level z_p such that $\text{GEV}(z_p) = 1 - p$

$$z_p = \text{GEV}^{-1}(1 - p)$$

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

Estimation

Two widely-used methods for estimation of $\boldsymbol{\theta} = (\mu, \sigma, \xi)^T$.

1. Numerical maximum likelihood
2. Probability weighted moments/L-moments

Arguments:

- Standard ML setting if $\xi > -.5$.
- $\heartsuit$ ML properties $\heartsuit$.
- *numerical* ML can be flaky for small sample sizes.
- PWM implicitly makes assumption $\xi < .5$.

Example: Hartford wind data

Data: annual maximum wind measurements at Hartford 1944-1983 (from `ismev` package).

Say we are interested in two quantities:

- $P(\text{annual max} > 70)$.
- 100-year return level.

R Demo

Example: Hartford wind data

Parameter Estimates

ML:

$\widehat{\mu} = 49.93, \widehat{\sigma} = 5.02, \widehat{\xi} = 0.004$

PWM:

$\widehat{\mu} = 50.01, \widehat{\sigma} = 5.24, \widehat{\xi} = -0.043$

Note: fundamental difference in estimates of ξ.

P(ann max > 70)

ML Pt Est: 0.019

PWM Pt Est: 0.015

Return Level Pt Est

ML Pt Est: 73.23

PWM Pt Est: 72.02

Uncertainty quantification: ML estimation

Aim: Provide uncertainty information about the parameter estimates.

The hessian of the likelihood surface can be numerically estiimated at the maximum likelihood estimate. Invert to estimate empirical information matrix.

Hartford Wind 95% CI's

μ: (48.21, 51.66)

σ: (3.77, 6.26)

ξ: (-0.19, 0.20)

Notes:

- parameter uncertainty can be done for PWM's, too.
- $\widehat{\xi}$ often found to have skewed distribution.

Uncertainty quantification: 100 year Rtn Level

Parameters not very interpretable. Better to provide uncertainty about a meaningful quantity e.g. 100-year return level.

Two methods:

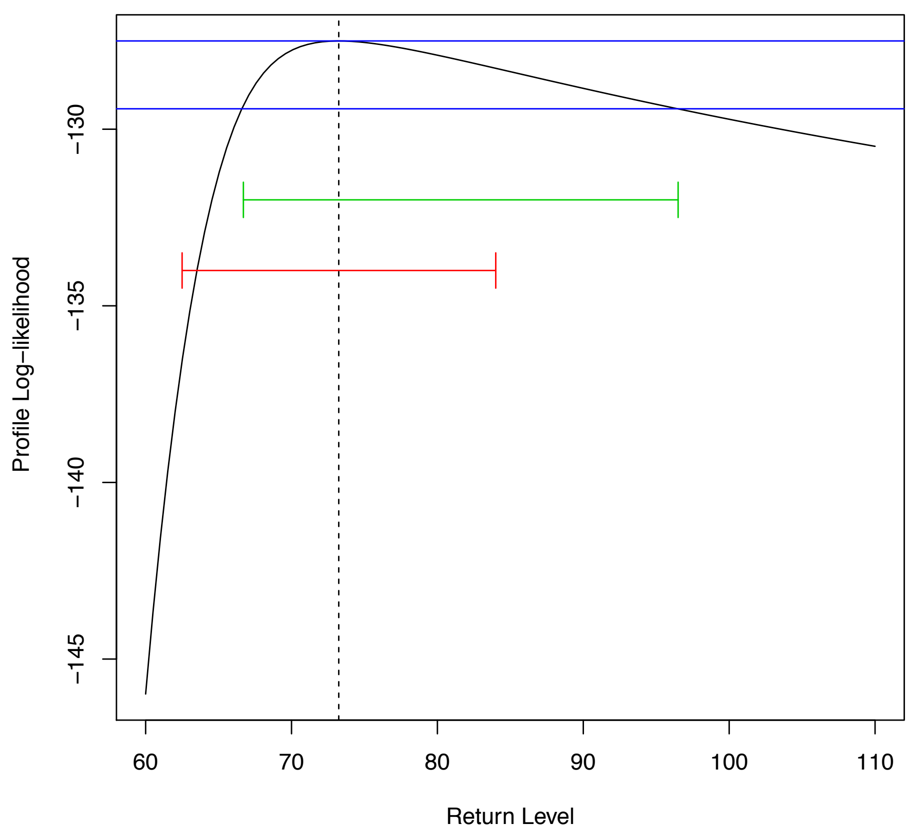
1. Delta-method (relies on asymptotic normality)
2. Profile likelihood

RDemo

Uncertainty quantification: 100 year Rtn Level

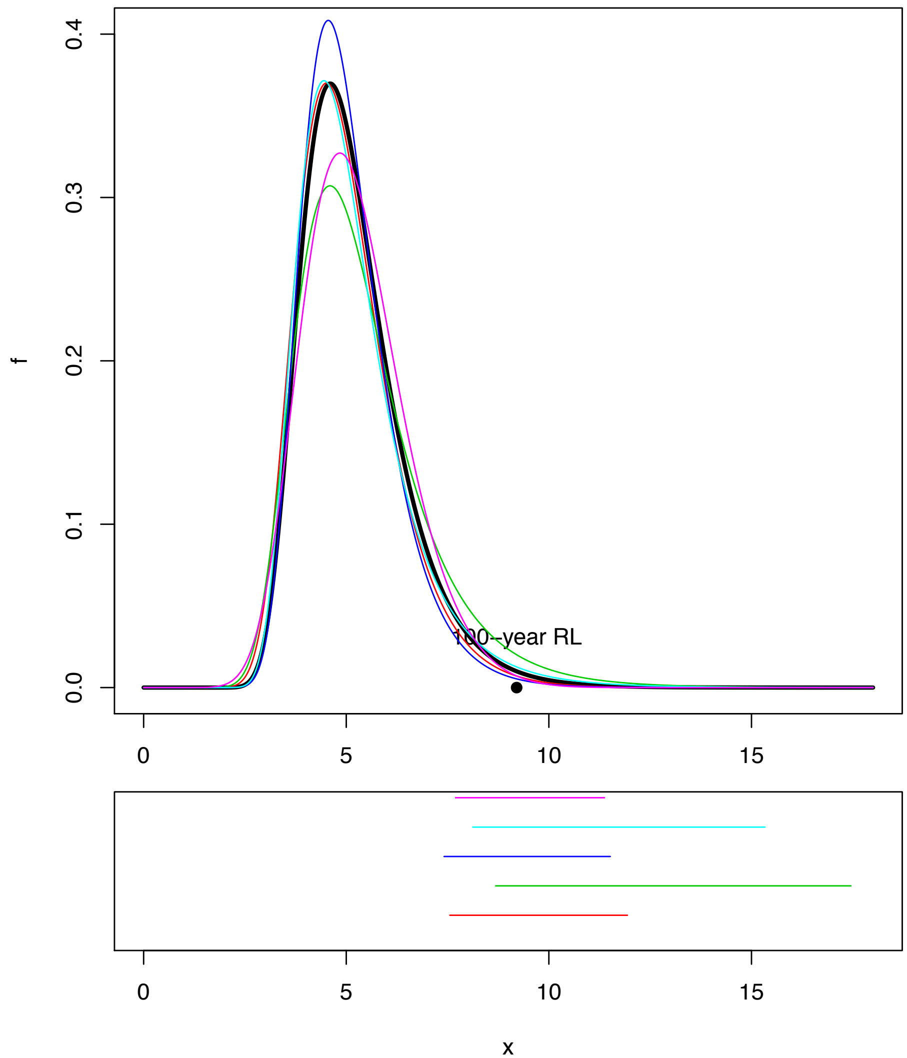
Parameters not very interpretable. Better to provide uncertainty about a meaningful quantity e.g. 100-year return level.

1. Delta-method: (66.7, 96.5) Note: error on handouts
2. Profile likelihood: (66.7, 96.5)



Simulation Study

Simulate 50 ‘years’ of 100 obs, estimate GEV, & 100-yr RL.



Coverage rate of 95% CI for 100-year RL: 0.93

Take-away messages from Part I

1. To estimate the tail, EVT uses only extreme observations.
2. The distribution of (renormalized) sample maxima converges to a max-stable distribution.
3. The three types of max-stable distributions are covered by the GEV distribution.
4. One approach to fitting extremes:
   (a) Choose a block length, extract data.
   (b) Fit a GEV (μ and σ are related to this block length).
   (c) Make inference, accounting for uncertainty.
   (d) End up modeling the distribution of the block (e.g., annual) maximum.
5. Tail parameter ξ is extremely important. Unfortunately, ξ is also very hard to estimate.
6. Not Gaussian!

Q: Is using only the block maxima wasteful of data?

GEV only *approximates* distn of block maximum

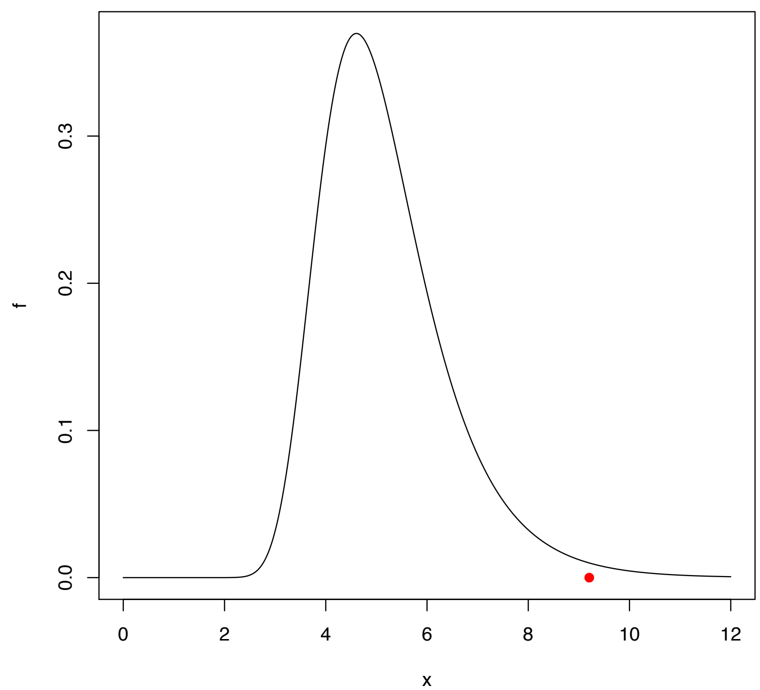
Let $\{X_t\}$ be an *iid* sequence of Exponential(1) RVs. Consider $M_{100} = \max(X_1, \ldots, X_{100})$.

Correct distribution:

$$P(M_{100} \le x) = [P(X_i \le x)]^{100} = [1 - \exp(-x)]^{100}.$$

$$f_{M_{100}}(x) = 100\,[1 - \exp(-x)]^{99} * \exp(-x).$$

$$\text{100-‘year’ return level: } = 9.205$$



example : maxima of normally distributed random variables

Home work simulation

- Generated random sample of length 100 from standard normal distribution and obtain maximum value (Repeated 40,000 times)
- Fit GEV distribution to sample of 40000 maxima
- Check if the estimate of ξ is around -0.1 but not zero

Penultimate approximation

Theory by Fisher and Tippett (1928)

For infinite block sizes, maxima of Gaussian belong to the Gumbel galaxy, but for finite samples sizes the estimated shape parameters belong to the Weibull galaxy (Von Mises conditions)

$$\xi_n = \frac{1}{hazard'(a_n)} \text{ with } hazard(x) = \frac{f(x)}{\overline{F}(x)} \text{ and } \overline{F}(x) = 1 - F(x)$$

Back to the Gaussian example

$$hazard(x) \approx x \text{ for large } x \text{ and } a_n \approx \sqrt{2 \log n}$$

Hence,

$$\boxed{\xi_n \approx \frac{1}{2 \log n} \text{ and } \frac{1}{2 \log 100} \approx -0.109}$$

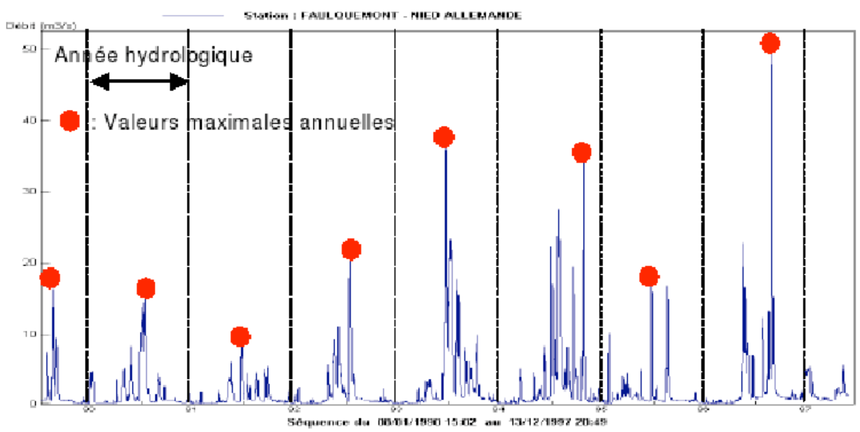
Lessons learned from this Gaussian example

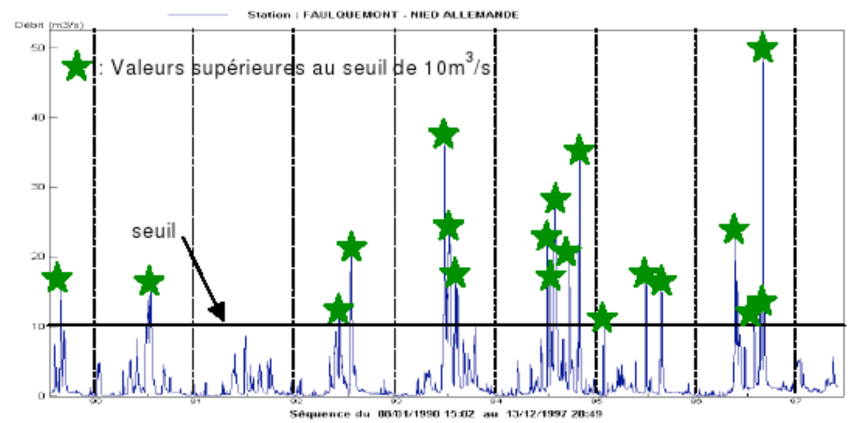
- The convergence towards max-stability can be very very slow
- The sign of the estimated shape parameter from a finite block size should not be over-interpreted
- Theoretical results exist to explain and quantity this phenomenon

Part II: Univariate threshold-exceedance approaches

- GPD and threshold stability.
- Inference.
- Threshold selection.
- Temporal dependence.
- Example.

Modeling exceedances





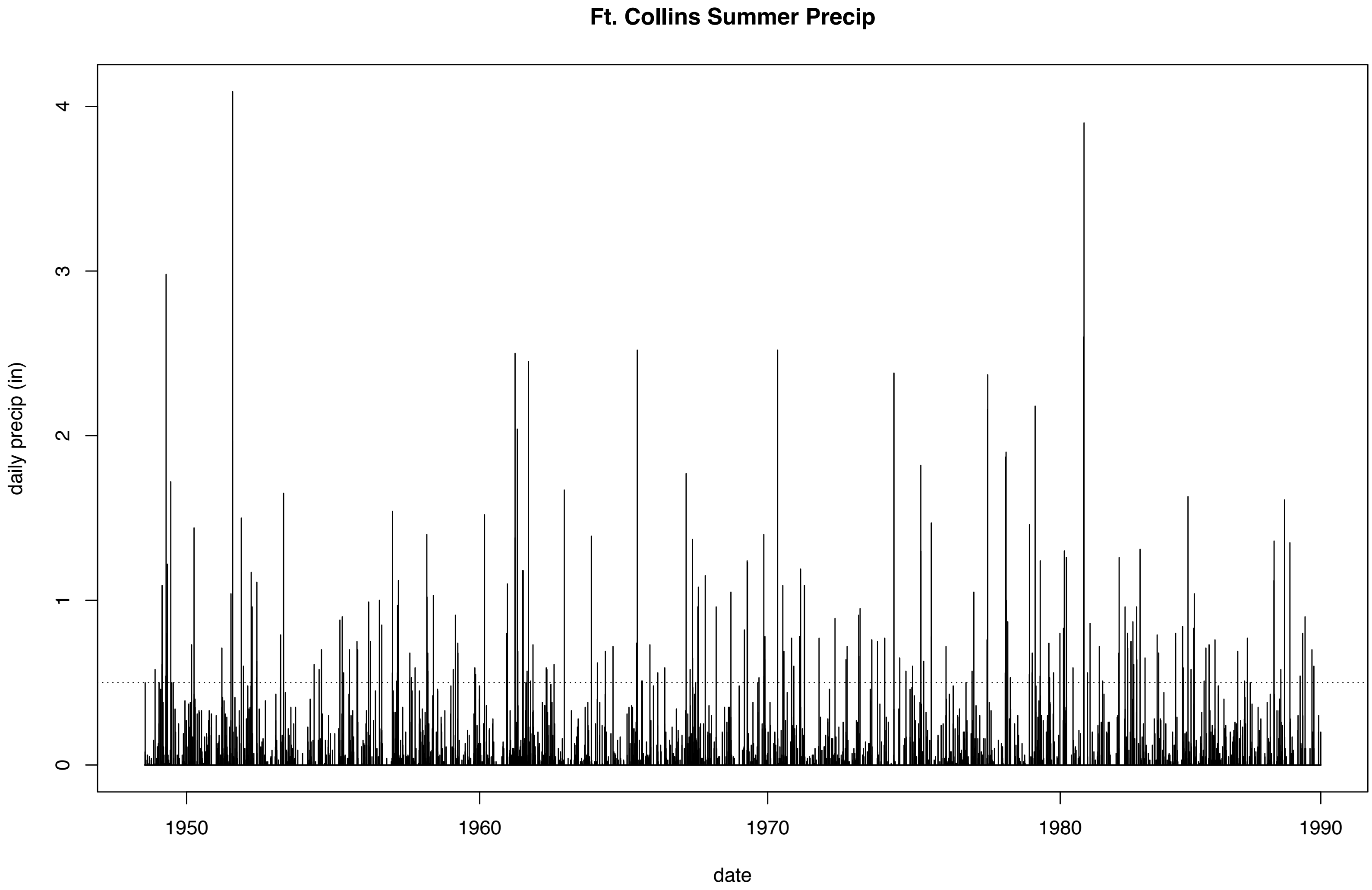
Generalized Pareto Distribution

the GPD is the right distribution for threshold exceedances.

$$P(X \leq z \mid X > u) = 1 - \left(1 + \frac{\xi(z-u)}{\psi_u}\right)^{-1/\xi}$$

In practice, choose threshold u, estimate ψ_u and ξ.

Fort Collins Data Revisited



GPD estimation of Fort Collins Precipitation

GEV	GPD
block size $=$ 214 days	$u = 0.5$ inches
$n = 42$	$n = 280$
$\widehat{\mu} = 1.11(0.09)$	
$\widehat{\sigma} = 0.46(0.07)$	$\widehat{\psi} = 0.39(0.04)$
$\widehat{\xi} = 0.31(0.18)$	$\widehat{\xi} = 0.19(0.07)$
$\widehat{r}_{100} = 5.8$ inches	$\widehat{r}_{100} = 6.1$ inches
95% CI: (3.5, 18.8)	95% CI: (4.3, 10.6)

- Parameter estimates and s.e's obtained via numerical maximum likelihood.
- GPD has lower standard error for ξ, lower estimate as well.
- GPD has narrower confidence interval.
- Have not yet discussed threshold selection procedure.

Threshold Stability

If GPD is a limiting distribution, it must be stable. How?

Assume X is GPD above some threshold u. Consider distribution of X for $X > u_1$, where $u_1 > u$.

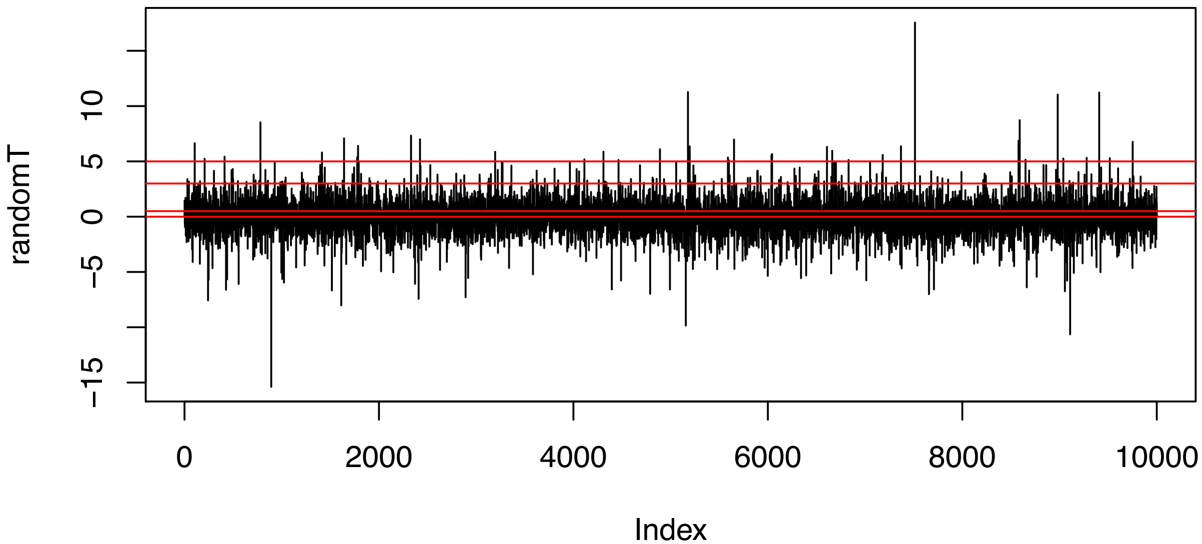
$$
\begin{aligned}
P(X > z \mid X > u_1) &= P(X > z \mid X > u_1, X > u) \\
&= \frac{P(X > z, X > u_1 \mid X > u)}{P(X > u_1 \mid X > u)} \\
&= \frac{\left(1 + \xi \frac{z-u}{\psi_u}\right)^{-1/\xi}}{\left(1 + \xi \frac{u_1-u}{\psi_u}\right)^{-1/\xi}} \\
&= \left(1 + \xi \frac{z - u_1}{\psi_{u_1}}\right)^{-1/\xi},
\end{aligned}
$$

where $\psi_{u_1} = \psi_u + \xi(u_1 - u)$, and ξ is unchanged.

Threshold Influence

Choice of threshold can strongly influence results.

Simulation of 10000 T-distributed RV's with 4 df.



True value of $\xi = 0.25$.
Estimation of a high quantile:

$$P(T > 13.03) = 0.0001$$

(essentially the 100-year return level if 100 obs per year).

Threshold Influence

Choice of threshold can strongly influence results.

Threshold	0	0.5	2	3	5	Truth
ξ	-.002(.010)	.062(.016)	.192(.048)	.207(.084)	.349(.224)	.25
Quantile	8.41(.30)	9.46(.52)	12.42(1.59)	12.75(2.05)	13.60 (3.32)	13.03
$P(T > u)$	.499	.322	.057	.019	.004	–

Threshold choice involves a tradeoff between bias and variance.

If threshold is too low, bias results because we are not far enough in the tail for EVT to apply.
If threshold is too high, estimators have high variance because there is not enough data to estimate parameters well.

Goal: find the lowest threshold u such that the tail behaves like a GPD above u.

Mean Residual Life

Assume that above a threshold u, the distribution is GPD.

$$P(X > z \mid X > u) = \left(1 + \frac{\xi(z-u)}{\psi_u}\right)^{-1/\xi}.$$

Consider the mean residual life: $E[X - u \mid X > u]$.

$$\begin{aligned} E[X - u \mid X > u] &= \int_u^{z_+} \left(1 + \frac{\xi(z-u)}{\psi_u}\right)^{-1/\xi} dz \\ &= \frac{\psi_u}{1-\xi} \text{ (via integration by sub).} \end{aligned}$$

$$\begin{aligned} \Rightarrow E[X - u_1 \mid X > u_1] &= \frac{\psi_{u_1}}{1-\xi}, \text{ for } u_1 > u \\ &= \frac{1}{1-\xi}\Big(\psi_u + \xi(u_1 - u)\Big) = cu_1 + d. \end{aligned}$$

That is, MRL is linear function of threshold when GPD holds.

Threshold selection: Mean residual life plots

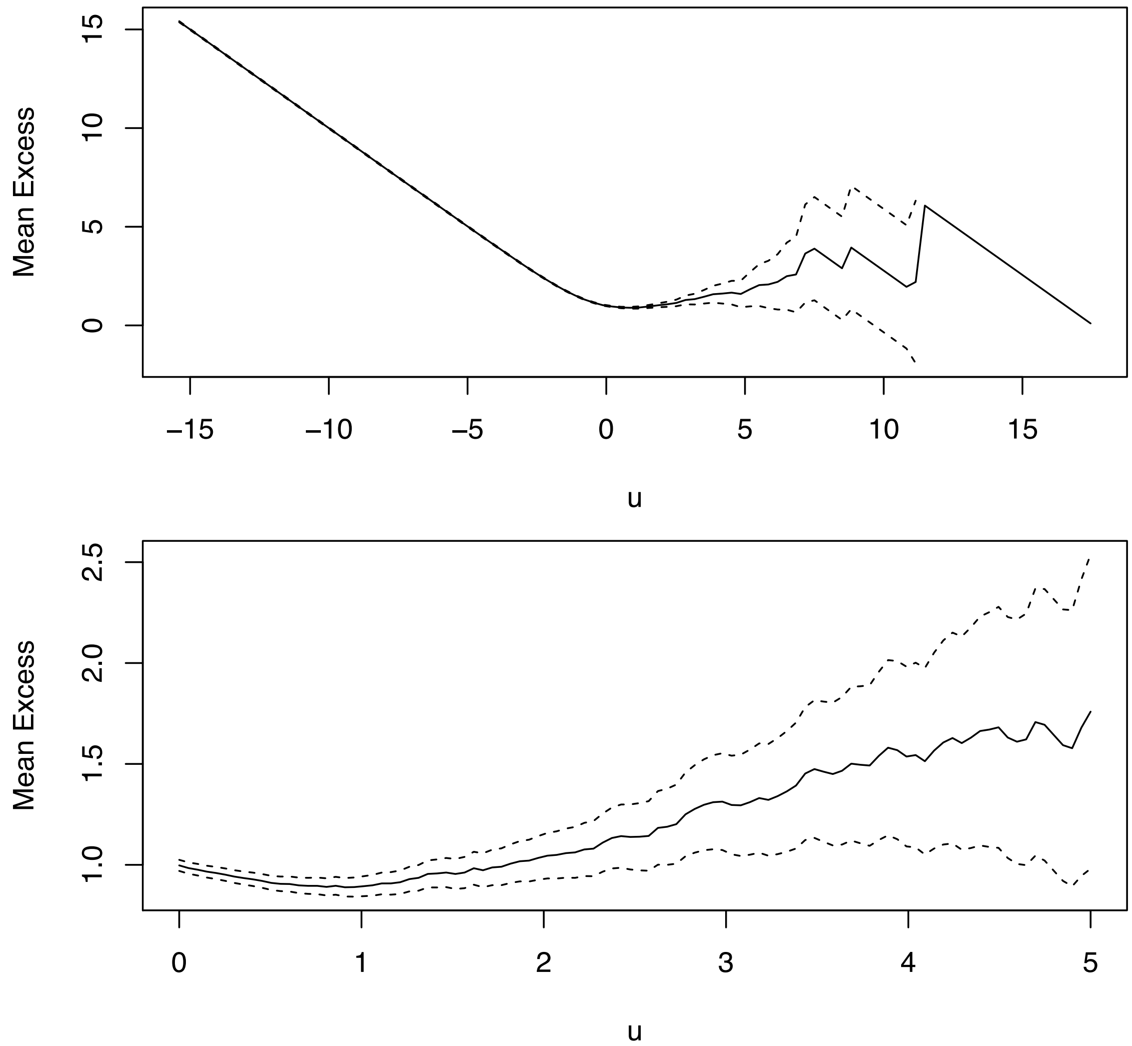
MRL should be linear where the tail behaves like a GPD.

Select a sequence of thresholds $u_1 < \ldots < u_n$. For each u_i, let $X_{(1)}, \ldots, X_{(N_{u_i})}$ represent the observations which exceed u_i. Calculate

$$eMRL(u_i) = \frac{1}{N} \sum_{j=1}^{N_{u_i}} X_j - u_i,$$

and plot $(u_i, eMRL(u_i))$, $i = 1, \ldots, n$.

Mean residual life plots for simulated T RV's

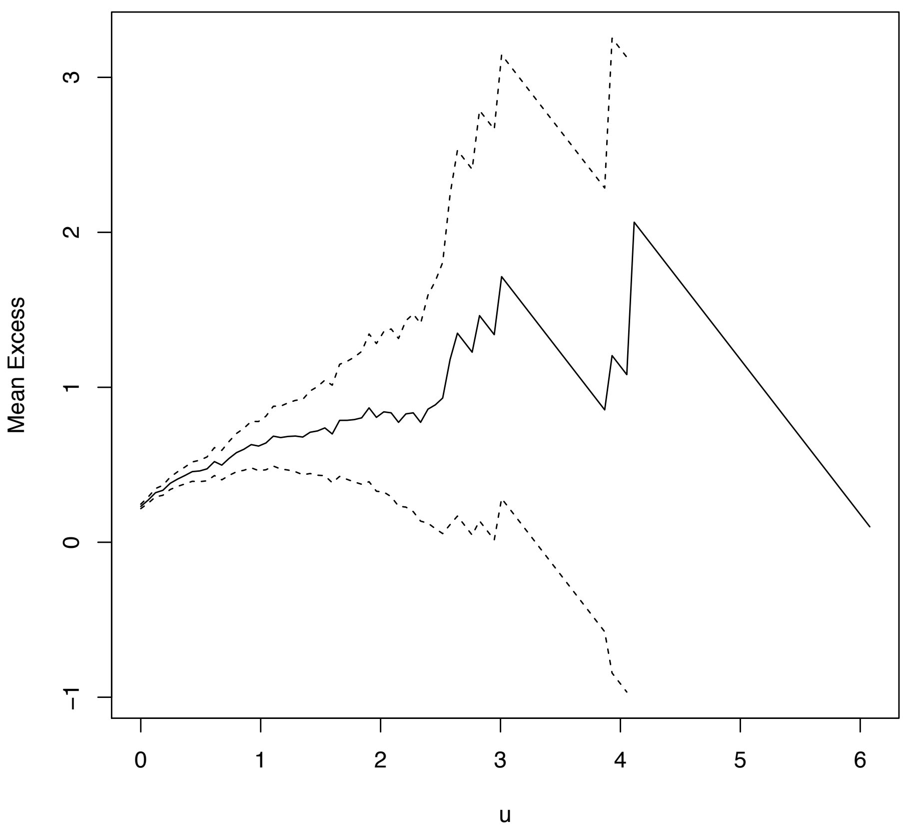


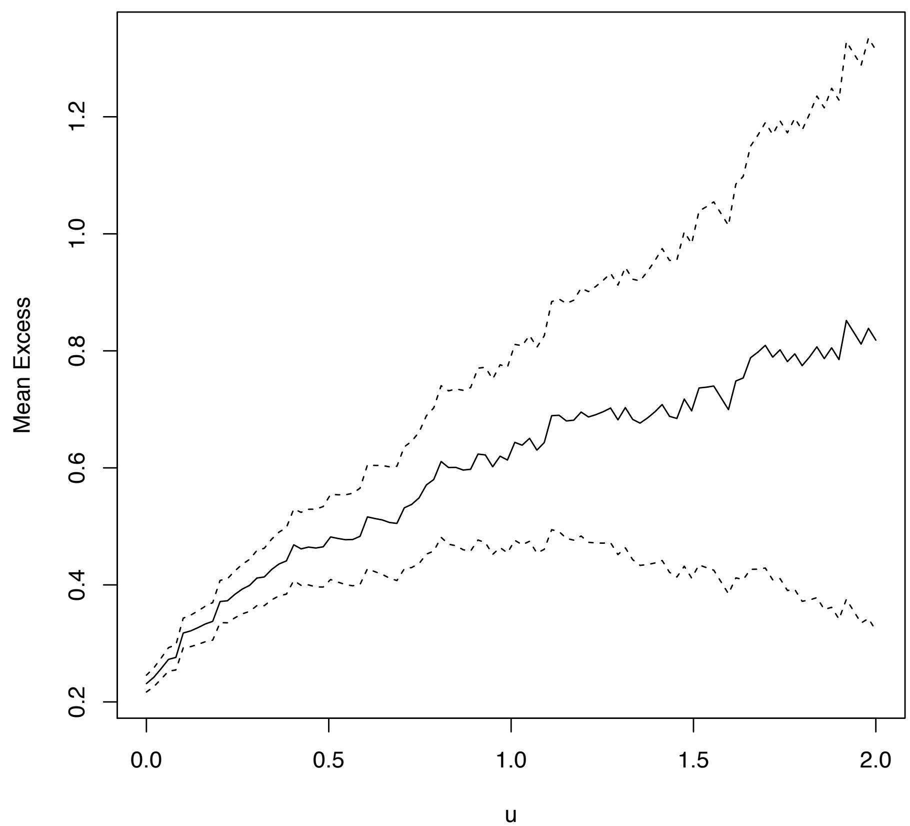
There is a bit of "art" to choosing a threshold based on MRL plots. And this is an easy example!

Threshold selection: Mean residual life plots

Recall: I set the threshold for the Fort Collins precipitation data at 0.5 inches.

MRL Plots for Fort Collins Data





Threshold Selection for Fort Collins Data

GPD fit to Fort Collins data with different thresholds.

	About Right	Too Low	Too High
u	0.5	0.05	2
N_u	280	1702	17
$\widehat{p}_{exc}$	0.025	0.156	0.0016
$\widehat{\psi}$	0.39 (0.04)	0.178 (0.007)	0.63 (0.238)
$\widehat{\xi}$	0.19 (0.07)	0.344 (0.032)	0.241 (0.30)
.9999 Qtile	4.42 (0.72)	6.02 (0.87)	4.44 (0.89)

Again see tradeoff between bias and variance.

Note: high variance of quantile estimate due to gradient term in delta method.

Threshold selection: Parameter plots

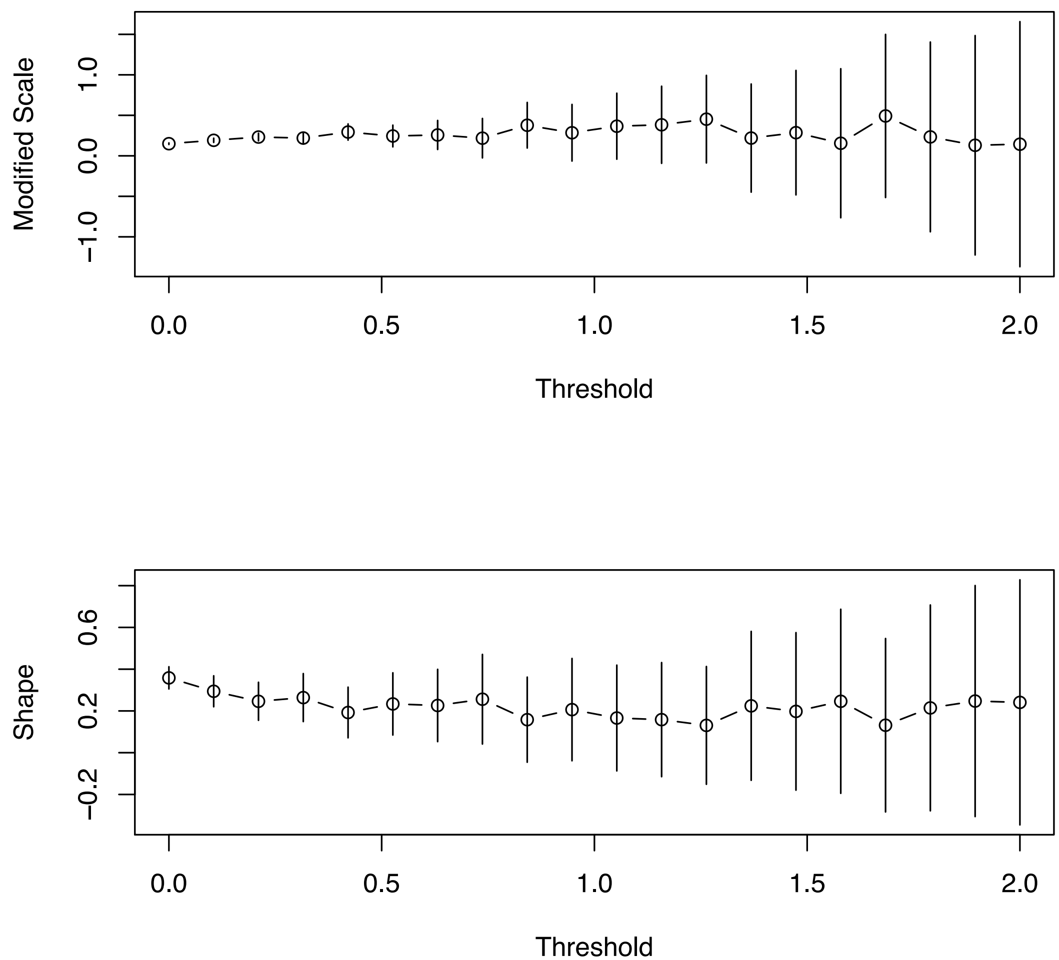
Assuming that X is GPD above u, recall that ξ is unchanged for higher thresholds.

Likewise, ψ can be rescaled so that it should be constant:

$$\psi_{u_1} = \psi_u + \xi(u_1 - u) \Rightarrow \psi_{u_1} - \xi u_1 = \psi_u - \xi u.$$

Threshold selection: Parameter plots

Parameter Plots for Fort Collins Data



Again, some “art” to interpreting these plots.

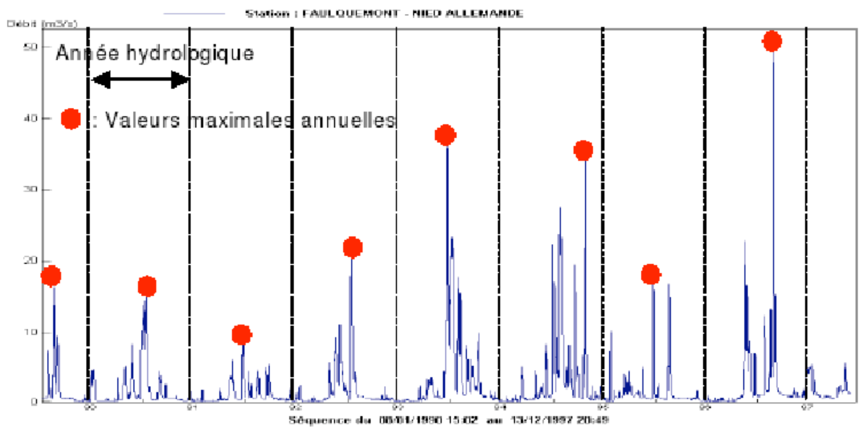
Take-away messages: Part II

2. GPD is the limiting distribution for threshold exceedances, exhibits threshold stability.
3. Threshold exceedance approaches (generally) allow the user to retain more data than block-maximum approaches, thereby reducing the uncertainty with parameter estimates, and consequently with return-level estimates.
4. Threshold selection is a classic bias/variance trade-off. Idea is to choose the threshold as low as possible such that the asymptotic theory is well-approximated.

Part III: Other topics in Univariate Extremes

- Extremes of stationary (but not independent) sequences.
- Extremes in the non-stationary case.

Modeling exceedances



Stationary Example

Let $\{Y_t\}$ be an iid sequence of RV's with cdf $F_Y(x) = \exp\left(-\frac{1}{(a+1)x}\right)$, where $a \in [0,1]$.

Let $\{X_t\} = \max(aY_{t-1}, Y_t)$. Notice:

$$\begin{aligned}
P(X_t \le x) &= P(aY_{t-1} \le x, Y_t \le x) \\
&\overset{\text{indep}}{=} P(aY_{t-1} \le x)P(Y_t \le x) \\
&= \exp\left(-\frac{a}{(a+1)x}\right)\exp\left(-\frac{1}{(a+1)x}\right) \\
&= \exp\left(-\frac{1}{x}\right).
\end{aligned}$$

- a introduces (positive) dependence into the sequence
- As $a \to 1$, extremes occur in pairs; $a = 0 \Rightarrow$ iid.
- The marginal is unit Fréchet for all a.

Q: Does dependence affect the extremes? *R Demo*

So What?

Is the GEV still right for our stationary sequence example?

If $M_n^* \sim GEV(\mu, \sigma, \xi)$, then

$$\begin{aligned}
P(M_n \le x) &= P(M_n^* \le x)^{1/(a+1)} \\
&= \left\{ \exp\left[- \left(1 + \xi \left(\frac{x-\mu}{\sigma} \right) \right)^{-1/\xi} \right] \right\}^{1/(a+1)} \\
&= \exp\left[\frac{-1}{a+1} \left(1 + \xi \left(\frac{x-\mu}{\sigma} \right) \right)^{-1/\xi} \right] \\
&= \exp\left[- \left(1 + \xi \left(\frac{x - (\mu + \sigma/\xi(1-(a+1)^{-\xi})}{\sigma(a+1)^{-\xi}} \right) \right)^{-1/\xi} \right.
\end{aligned}$$

So $M_n \sim GEV(\mu', \sigma', \xi)$.

Whew! This is BIG NEWS! But is it true in general or just for our example?

Extremes of Stationary Sequences

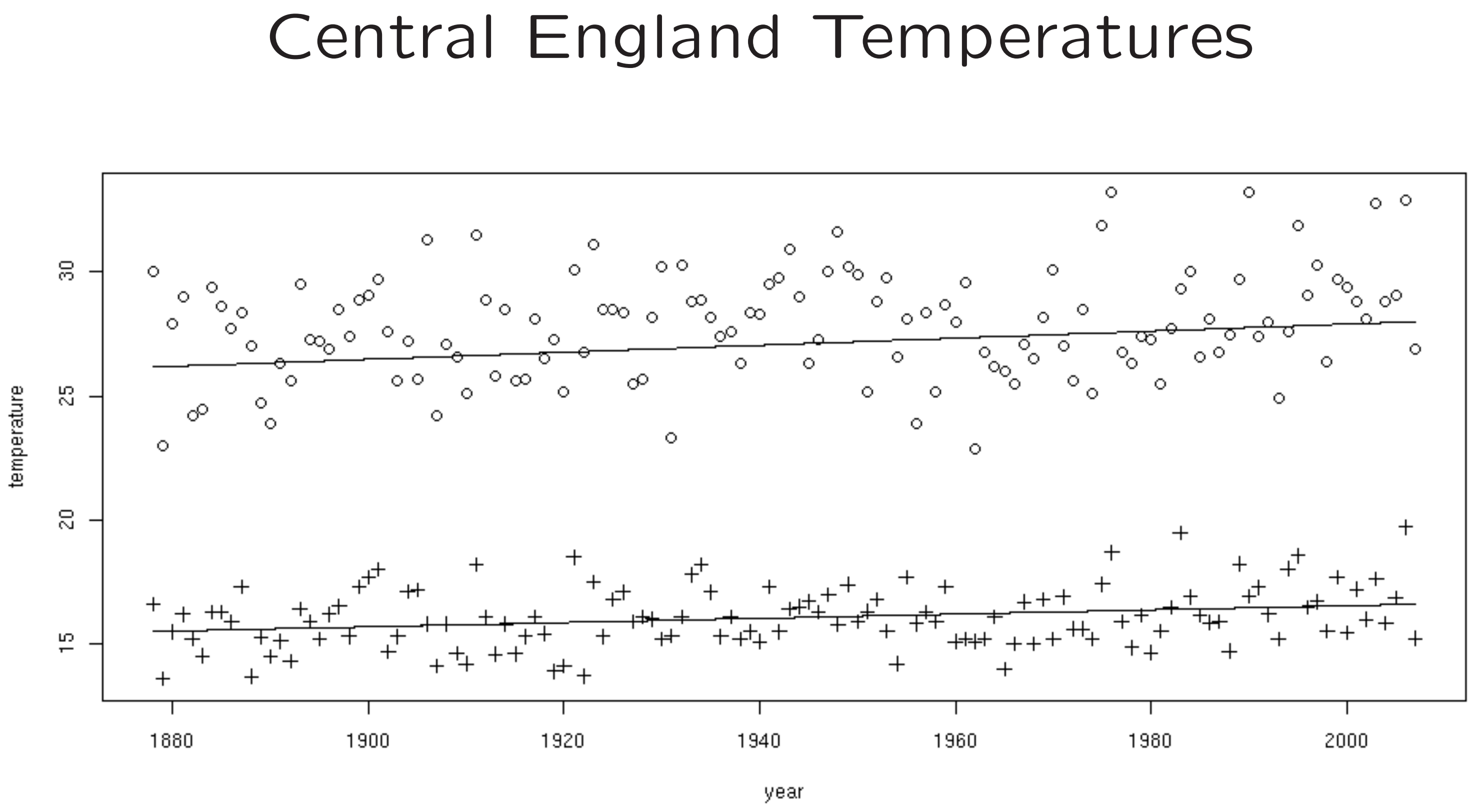
Q: Is the GEV still the limiting distribution for block maxima of a stationary (but not independent) sequence $\{X_t\}$?
A: Yes, so long as (relatively weak) mixing conditions hold. (Leadbetter et al., 1983)

What does this mean for inference?

Block maximum approach: GEV still correct for marginal. Since *block maximum* data likely have negligible dependence, proceed as usual.

Threshold exceedance approach: GPD is correct distribution for the marginal. If extremes occur in clusters, estimation affected as likelihood assumes independence of threshold exceedances.

Non-stationary Data



Q: What do you do when there is a trend or other behavior which you want/need to capture?

A: Allow the *parameters* of the EVD to vary with time or other covariates.

Non-stationary Example: England Temperature

$$P(M_n(t) \leq x) = \exp\left\{ - \left[1 + \xi(t) \left(\frac{x - \mu(t)}{\sigma(t)} \right) \right]^{-1/\xi} \right\}.$$

Here we let:

$$\begin{aligned} \mu(t) &= a + bt \\ \sigma(t) &= \sigma \\ \xi(t) &= \xi \end{aligned}$$

Estimates:

$$\begin{aligned} \widehat{a} &= 26.17 \\ \widehat{b} &= 0.0142 \\ \widehat{\sigma} &= 2.04 \\ \widehat{\xi} &= -0.27 \end{aligned}$$

Estimated annual max temperature exceedance probabilities:
Stationary Model: $P(M_n > 31.5) = 0.01.$
Nonstationary Model: $P(M_n(2007) > 31.5) = 0.10$
$P(M_n(2030) > 31.5) = 0.13.$

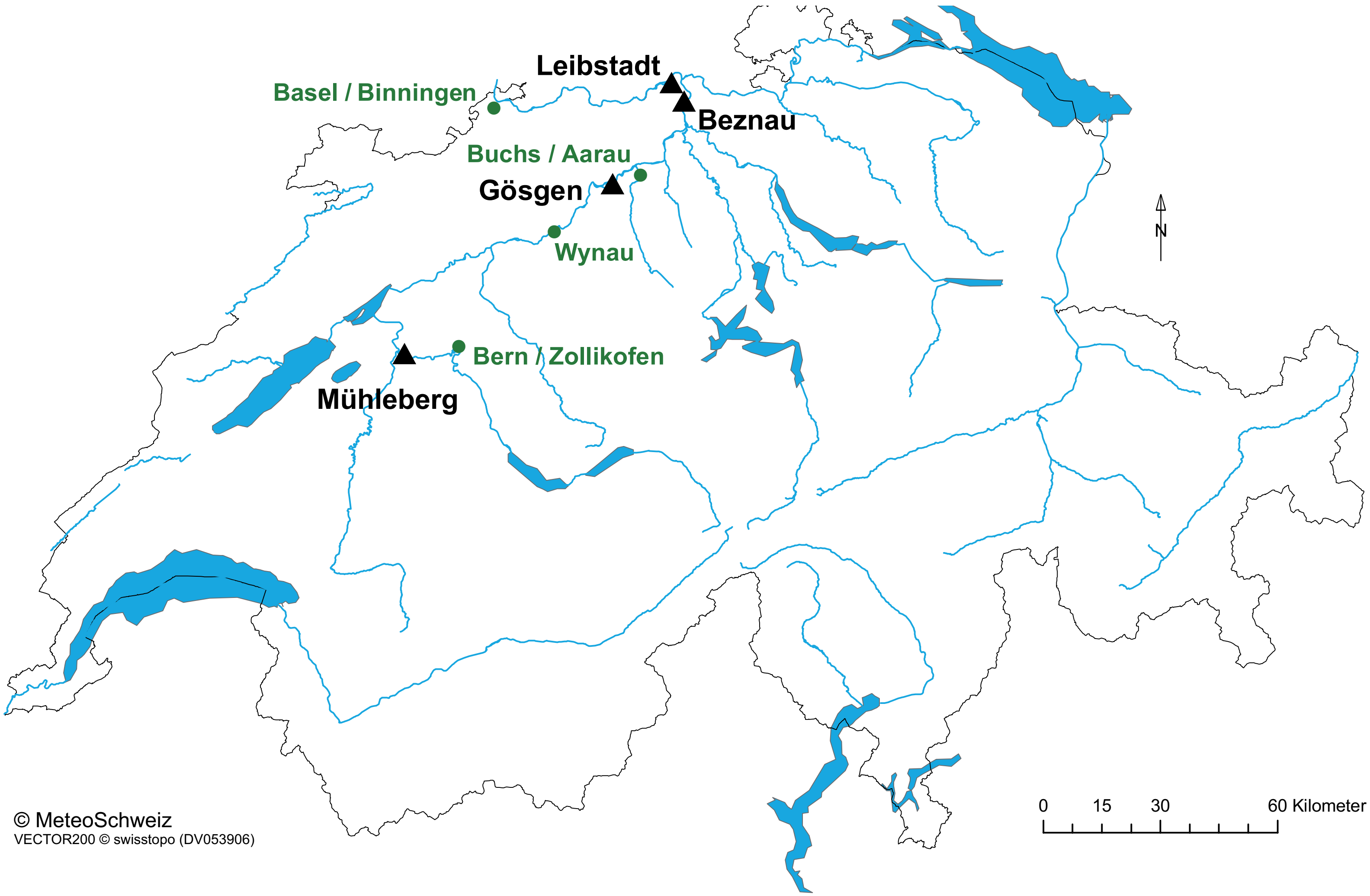
Nuclear power safety

- Fukushima ⇒ nuclear power safety concerns worldwide
- Swiss nuclear regulator asked for (re-)assessment of vulnerability of the four nuclear plants to
  - high and low air temperatures
  - high and low river water temperatures
  - high winds (and tornados)
  - intense rainfall, snowload, lightning strikes,
  - earthquakes and any tsunamis are dealt with separately!
- Task: estimate quantiles for probabilities 10^{-4} per year (and 10^{-7} for high winds), and give their uncertainties

Nuclear power safety

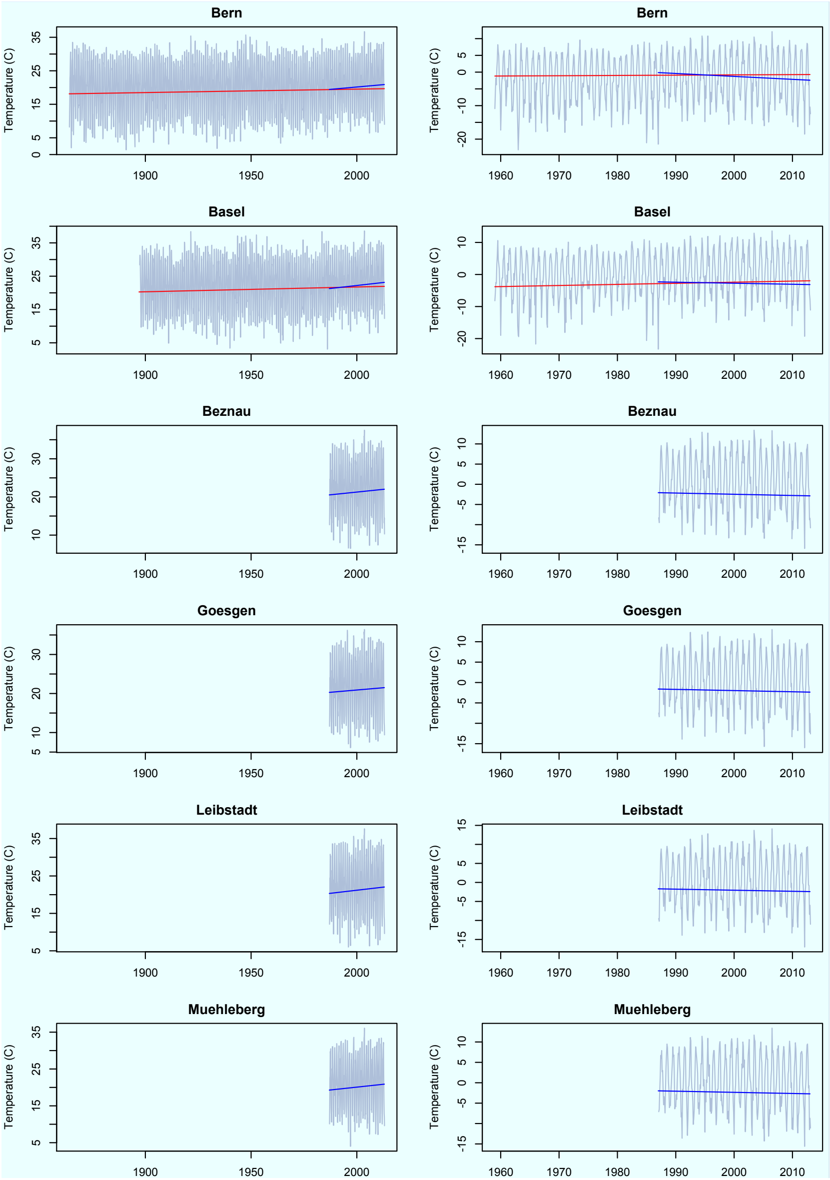
- Fukushima $\Rightarrow$ nuclear power safety concerns worldwide
- Swiss nuclear regulator asked for (re-)assessment of vulnerability of the four nuclear plants to
  - high and low air temperatures
  - high and low river water temperatures
  - high winds (and tornados)
  - intense rainfall, snowload, lightning strikes,
  - earthquakes and any tsunamis are dealt with separately!
- Task: estimate quantiles for probabilities 10^{-4} per year (and 10^{-7} for high winds), and give their uncertainties
  - based on 25 years of data or so at the plants themselves, and (at very most, and only for comparison) 150 years of data nearby

Swiss nuclear plants



Muhleberg

Air temperature maxima and minima



And one we took . . .

- Using annual maxima loses too much information
- Monthly maxima gives 12 (more likely ~ 6) 'independent' observations/year
- Allow for seasonality/trend by fitting GEV(μ_t, σ, ξ), with (e.g.)

$$\mu_t = \alpha_0 + \alpha_1 t + \sum_{m=1}^{M} \{\beta_m \sin(2\pi mt/365) + \gamma_m \cos(2\pi mt/365)\}$$

  with t the day on which the monthly maximum appears.
- Choose among models with/without trend, different values of M, using AIC and BIC, seeking a single compromise model for each variable
- If we're unlucky, then we also need σ_t and (if very unlucky) ξ_t
- Estimates of α_1 are totally unrealistic with short time series: $\widehat{\alpha}_1 \approx 7.5^\circ$/century for Leibstadt air temperature maxima
- Many variables have $\widehat{\xi} < 0$, so are bounded (phew!)
- End up with 'stable reference model' for standard year (1998), from which we extrapolate forward

Data analysis

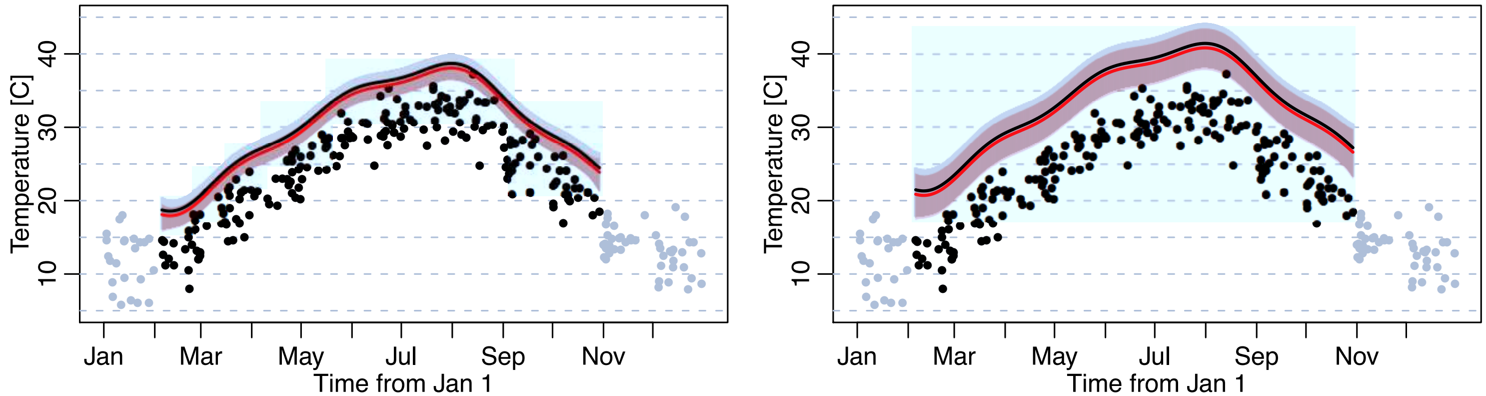
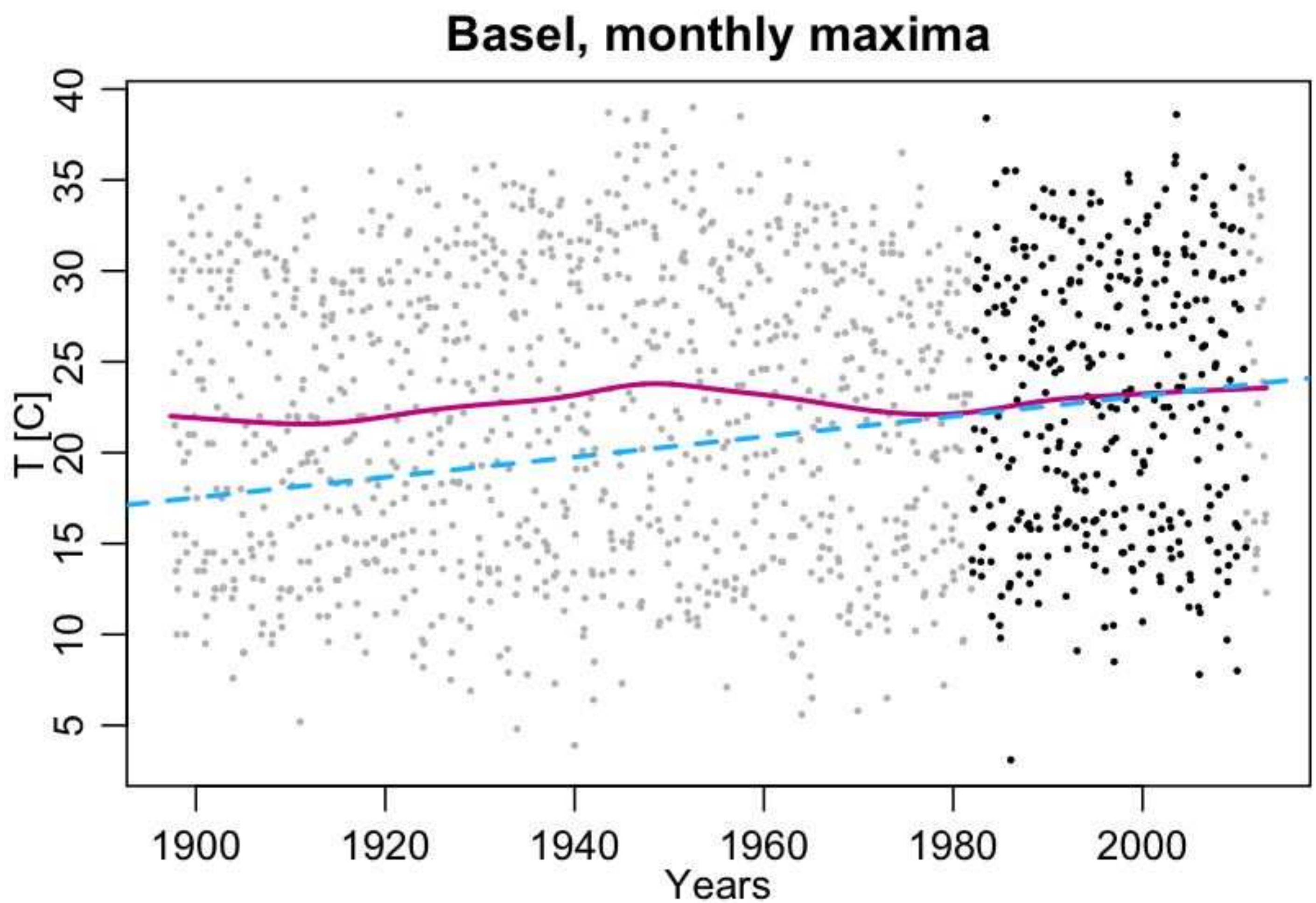


Figure 9: Seasonal 50- and 10000 year return levels (red and black, respectively) as a function of time in 1998 (left, coinciding with the reference model in Figure 7) and in 2050 (right). 95% confidence intervals from parametric bootstrap are shown as pink and light grey bands.

- Fitted GEV to monthly maxima for winter/summer seasons, allowing for monthly variation in location and (linear!) time trend
- Estimated shape parameter $\widehat{\xi} < 0$ implies upper bound on maximal temperature
- Attempt to allow for uncertainty due to
  - parameter estimation
  - number of observations contributing to maximum $(30 \neq \infty)$
  - stochastic variation of future events
  - changes of instrumentation (especially for winds)



- Back-extrapolation fails. Why should forward extrapolation succeed?

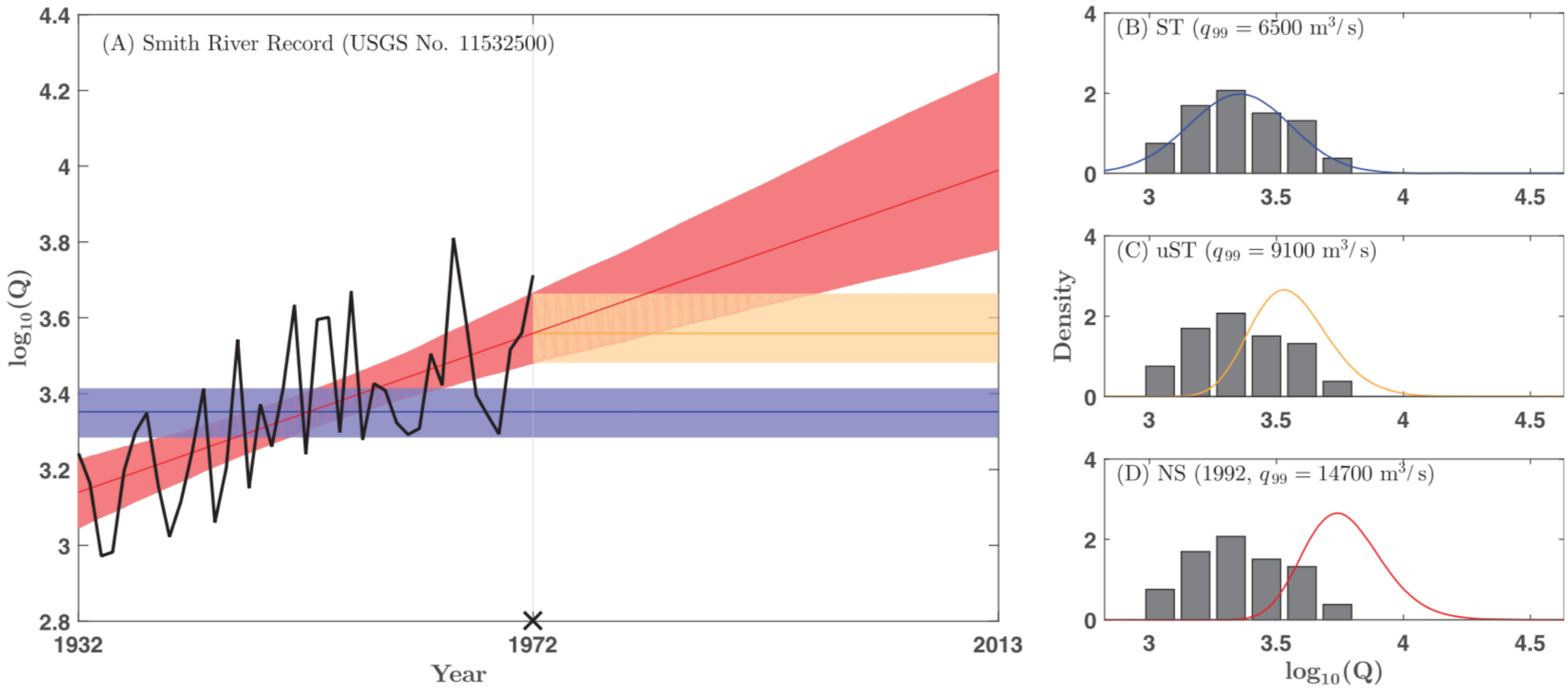


Figure 2. (a) Maximum a posteriori (MAP) estimate of the LPIII mean under the ST (*blue* line), uST (*gold* line), and NS (*red* line) models inferred from the Smith River fitting period **Q** (*black* line). The colored shading represents the respective 95% credible intervals of the LPIII mean, and the *black* cross denotes the end of the fitting period. (b–d) Predictions of out-of-sample density under the ST (*blue* line), uST (*gold* line), and NS (*red* line) models derived from the MAP parameter estimates. The *black* histograms represent the empirical density of the fitting period. Notice that Figures 2c and 2d show predictions under the uST and NS models moving *away* from the observed density, and the 95% credible intervals are wider under the NS and uST models relative to the ST model.

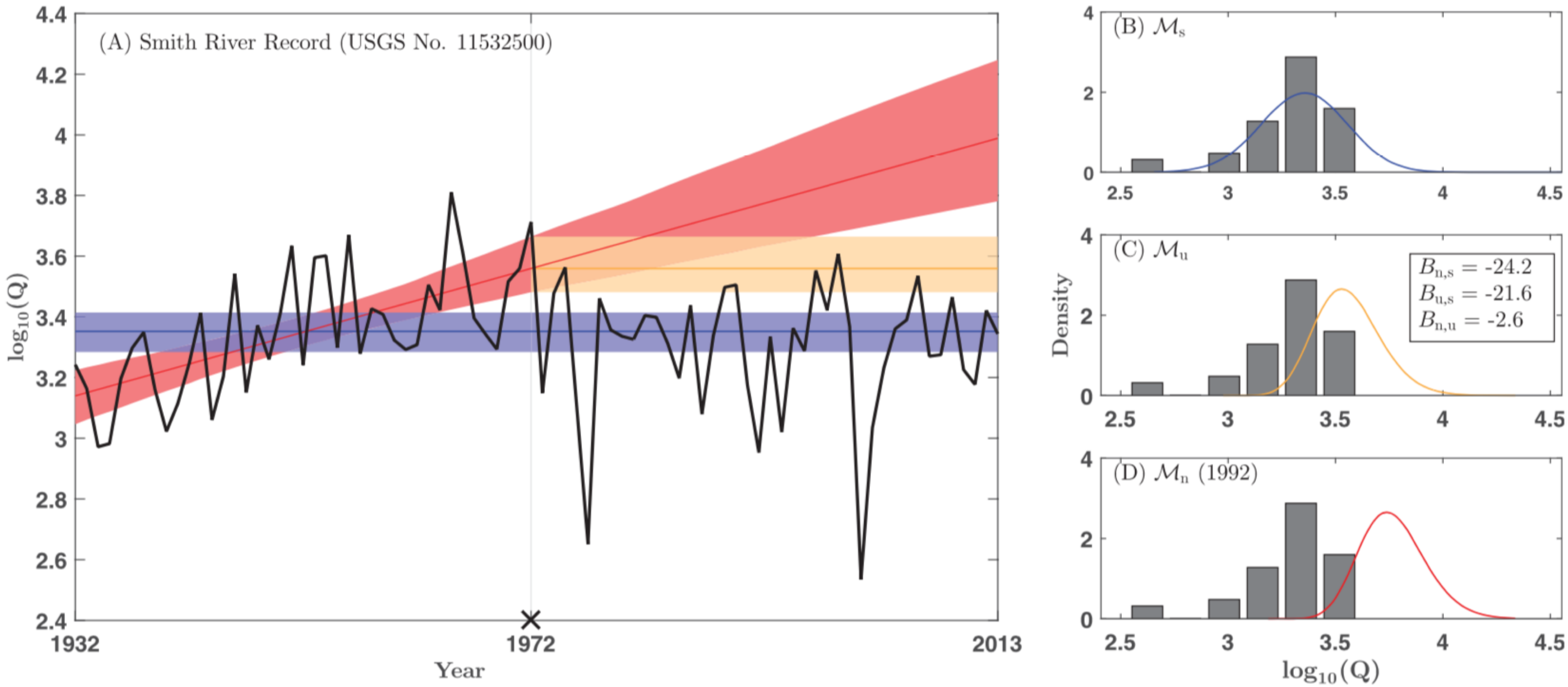
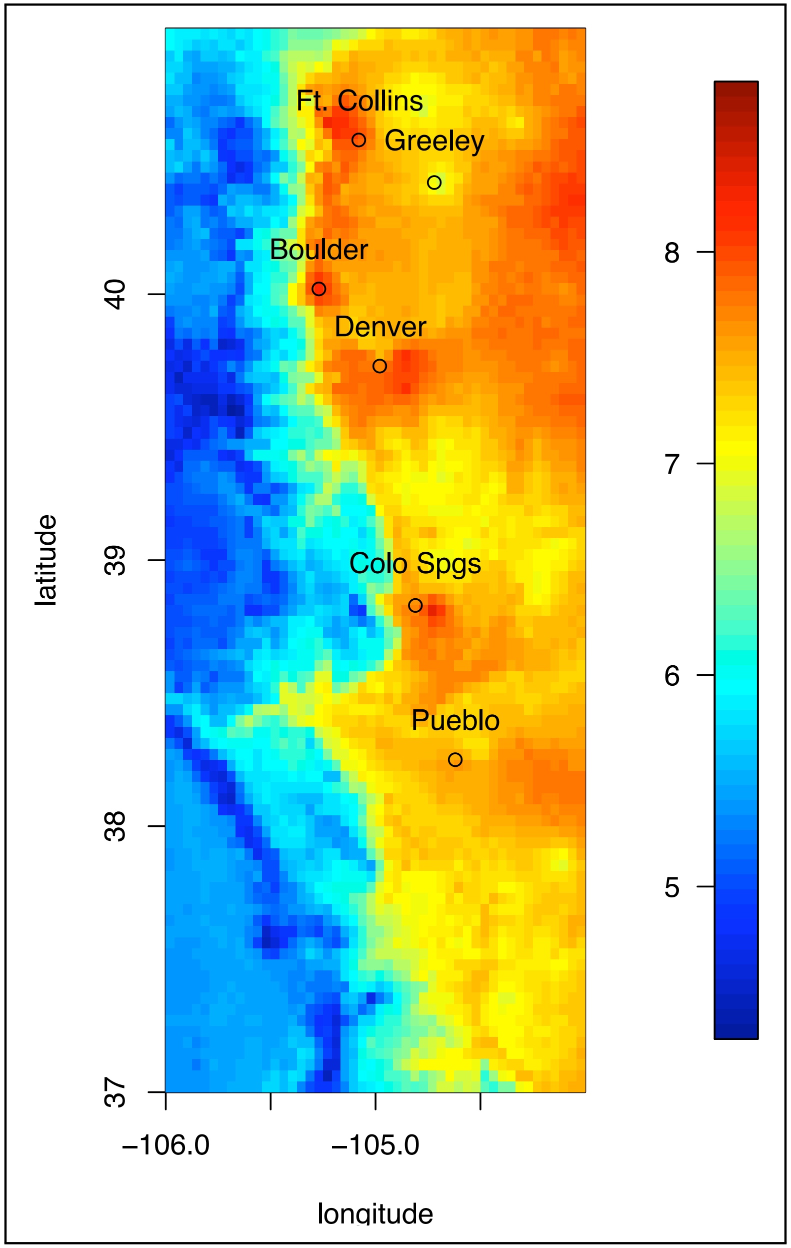


Figure 8. (a) The MAP estimate of the LPIII mean under $\mathcal{M}_s$ (*blue* line), $\mathcal{M}_u$ (*gold* line), and $\mathcal{M}_n$ (*red* line) shown over the full record length (*black* line). The colored shading represents the respective 95% credible intervals of the the LPIII mean. (b–d) Predictions of out-of-sample density under $\mathcal{M}_s$ (*blue* line), $\mathcal{M}_u$ (*gold* line), and $\mathcal{M}_n$ (*red* line) derived from the MAP parameter estimate. The *black* histograms show the empirical density in the evaluation period. For the Smith River record, $\mathcal{M}_s$ most accurately predicted the out-of-sample data, which is reflected by $B_{n,s}$ and $B_{u,s}$ shown in Figure 8c.

Return levels posterior mean



Our main assumptions

- Process layer : The scale and shape GPD parameters $(\xi(x), \sigma(x))$ are random fields and result from an unobservable latent spatial process
- Conditional independence : precipitation are independent given the GPD parameters

Our main variable change

$$\sigma(x) = \exp(\phi(x))$$

Our three levels

A) **Data layer** $:=$ pr(data|process, parameters) $=$

$$\mathrm{pr}_\theta\{\mathbf{R(x_i)} - u > y | \mathbf{R(x_i)} > u\} = \left(1 + \frac{\xi_i\, y}{\exp \phi_i}\right)^{-1/\xi_i}$$

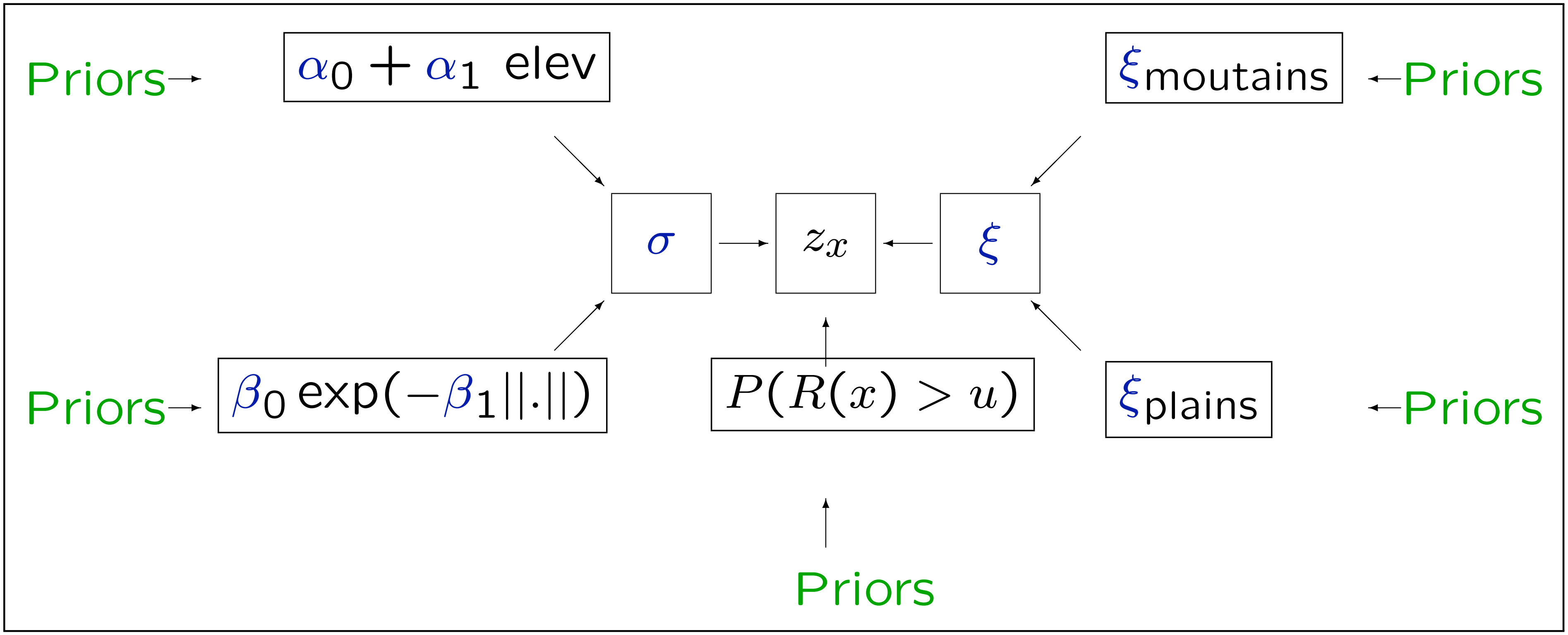
B) **Process layer :=** pr(process|parameters) :

$$\text{e.g. } \phi_i = \alpha_0 + \alpha_1 \times \text{elevation}_i + \text{MVN}\left(0, \beta_0 \exp(-\beta_1 ||x_k - x_j||)\right)$$

$$\text{and } \xi_i \quad = \quad \begin{cases} \xi_{\text{moutains}}, \text{ if } x_i \in \text{mountains} \\ \xi_{\text{plains}}, \text{ if } x_i \in \text{plains} \end{cases}$$

C) **Parameters layer (priors) :=** pr(parameters) :
e.g. ($\xi_{\text{moutains}}, \xi_{\text{plains}}$) Gaussian distribution with non-informative mean and variance

Bayesian hierarchical modeling

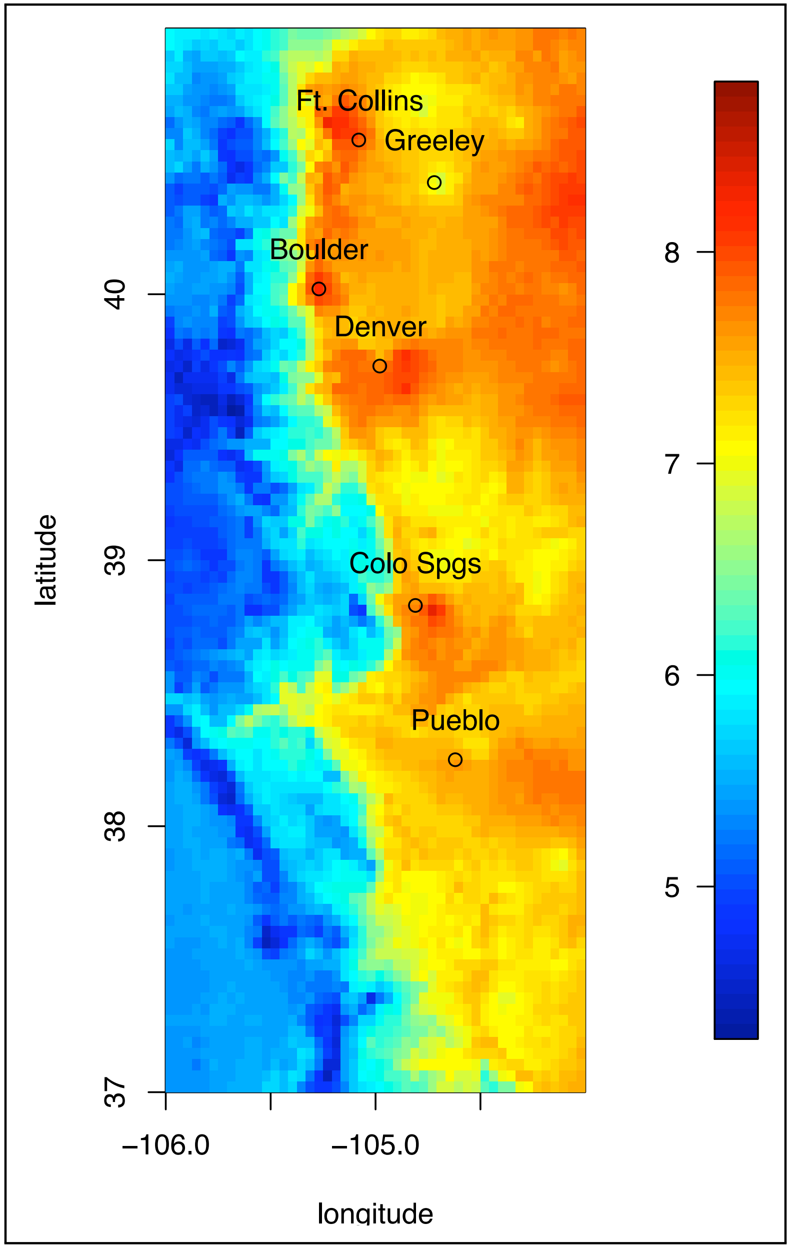


Model selection

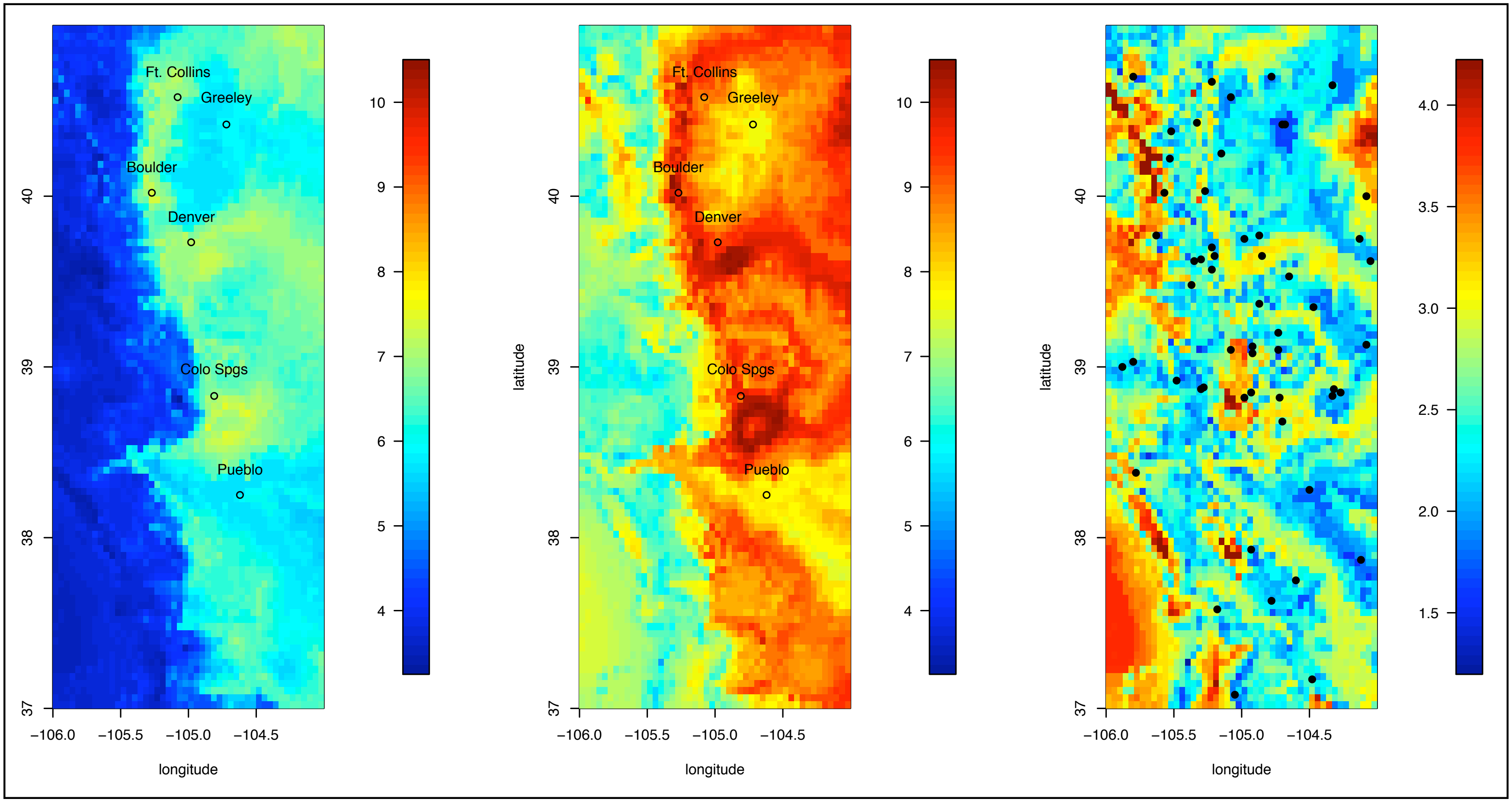
Baseline model		$\bar{D}$	p_D	*DIC*
Model 0:	$\phi = \phi$ $\xi = \xi$	73,595.5	2.0	73,597.2
Models in latitude/longitude space		$\bar{D}$	p_D	*DIC*
Model 1:	$\phi = \alpha_0 + \epsilon_\phi$ $\xi = \xi$	73,442.0	40.9	73,482.9
Model 2:	$\phi = \alpha_0 + \alpha_1(\text{msp}) + \epsilon_\phi$ $\xi = \xi$	73,441.6	40.8	73,482.4
Model 3:	$\phi = \alpha_0 + \alpha_1(\text{elev}) + \epsilon_\phi$ $\xi = \xi$	73,443.0	35.5	73,478.5
Model 4:	$\phi = \alpha_0 + \alpha_1(\text{elev}) + \alpha_2(\text{msp}) + \epsilon_\phi$ $\xi = \xi$	73,443.7	35.0	73,478.6
Models in climate space		$\bar{D}$	p_D	*DIC*
Model 5:	$\phi = \alpha_0 + \epsilon_\phi$ $\xi = \xi$	73,437.1	30.4	73,467.5
Model 6:	$\phi = \alpha_0 + \alpha_1(\text{elev}) + \epsilon_\phi$ $\xi = \xi$	73,438.8	28.3	73,467.1
Model 7:	$\phi = \alpha_0 + \epsilon_\phi$ $\xi = \xi_{\text{mtn}}, \xi_{\text{plains}}$	**73,437.5**	**28.8**	**73,466.3**
Model 8:	$\phi = \alpha_0 + \alpha_1(\text{elev}) + \epsilon_\phi$ $\xi = \xi_{\text{mtn}}, \xi_{\text{plains}}$	73,436.7	30.3	73,467.0
Model 9:	$\phi = \alpha_0 + \epsilon_\phi$ $\xi = \xi + \epsilon_\xi$	73,433.9	54.6	73,488.5

NOTE: Models in the climate space had better scores than models in the longitude/latitude space. $\epsilon_\cdot \sim \text{MVN}(0, \Sigma)$, where $[\sigma]_{i,j} = \beta_{\cdot,0} \exp(-\beta_{\cdot,1} \|\mathbf{x}_i - \mathbf{x}_j\|)$.

Return levels posterior mean



Posterior quantiles of return levels (.025, .975)



Pearson Type III

$$P(x) = \frac{1}{\beta\Gamma(p)}\left(\frac{x-\alpha}{\beta}\right)^{p-1}\exp\left(-\frac{x-\alpha}{\beta}\right).$$

Gamma-Pareto

$$f(x) = \frac{1}{\sigma} h_{\xi}\left(\frac{x}{\sigma}\right) g\left\{H_{\xi}\left(\frac{x}{\sigma}\right)\right\}$$

Gamma-Pareto

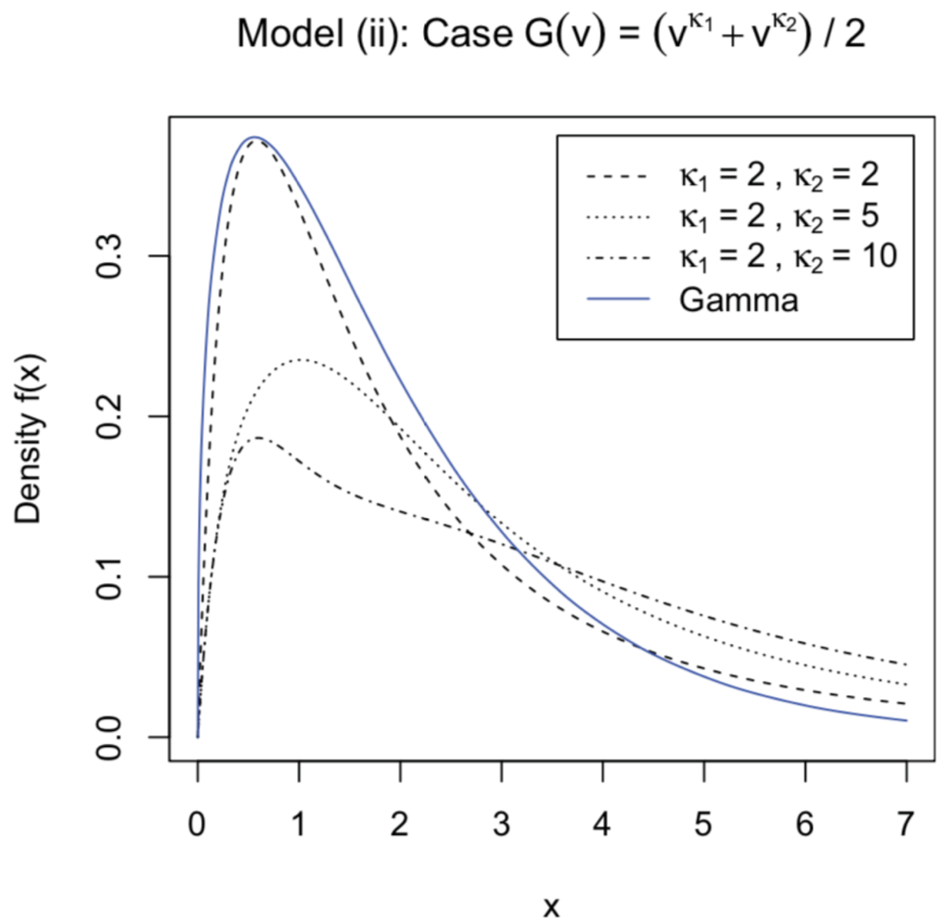


Figure 2. Density function corresponding to Model (6) combined with $G(v) = (v^{\kappa_1} + v^{\kappa_2})/2$ (a special case of Model (ii) with $p = 0.5$), for $\sigma = 1$, $\xi = 0.5$ and parameters $\kappa_1 = 2$, and $\kappa_2 = 2, 5, 10$ (dashed-dotted, dotted, dashed black curves, respectively). The solid blue curve represents a gamma density with parameters (1.4, 1.4).

Gamma-Pareto

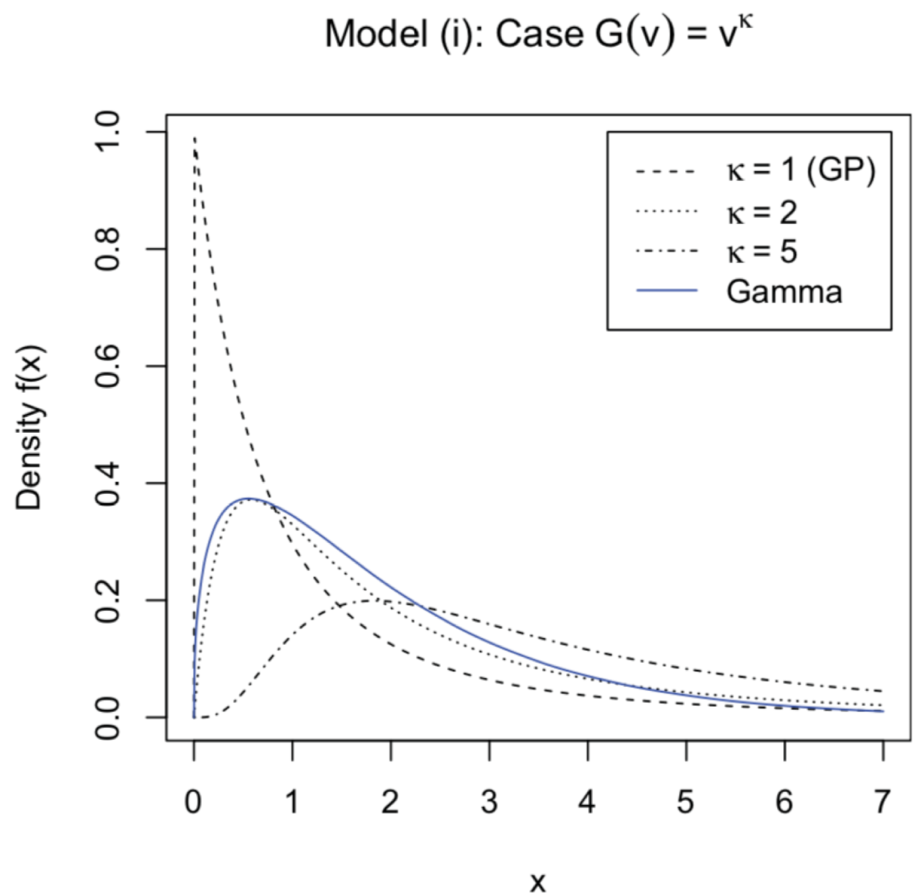


Figure 1. Density function corresponding to Model (6) with $G(v) = v^{\kappa}$, for $\sigma = 1$, $\xi = 0.5$ and lower tail shape parameters $\kappa = 1, 2, 5$ (dashed, dotted, dashed-dotted black curves, respectively). The case $\kappa = 1$ corresponds to the exact GP density. The solid blue curve represents a gamma density with parameters (1.4, 1.4).

Take-Away Messages Part III

- Stationary sequences
  - GEV is still correct limiting distribution for most stationary sequences.
  - Inference for block maximum methods unchanged.
  - Threshold exceedance approaches need to account for clustering of exceedances.
- Extremes of nonstationary sequences can be modeled using a regression approach on the parameters of the EVD.

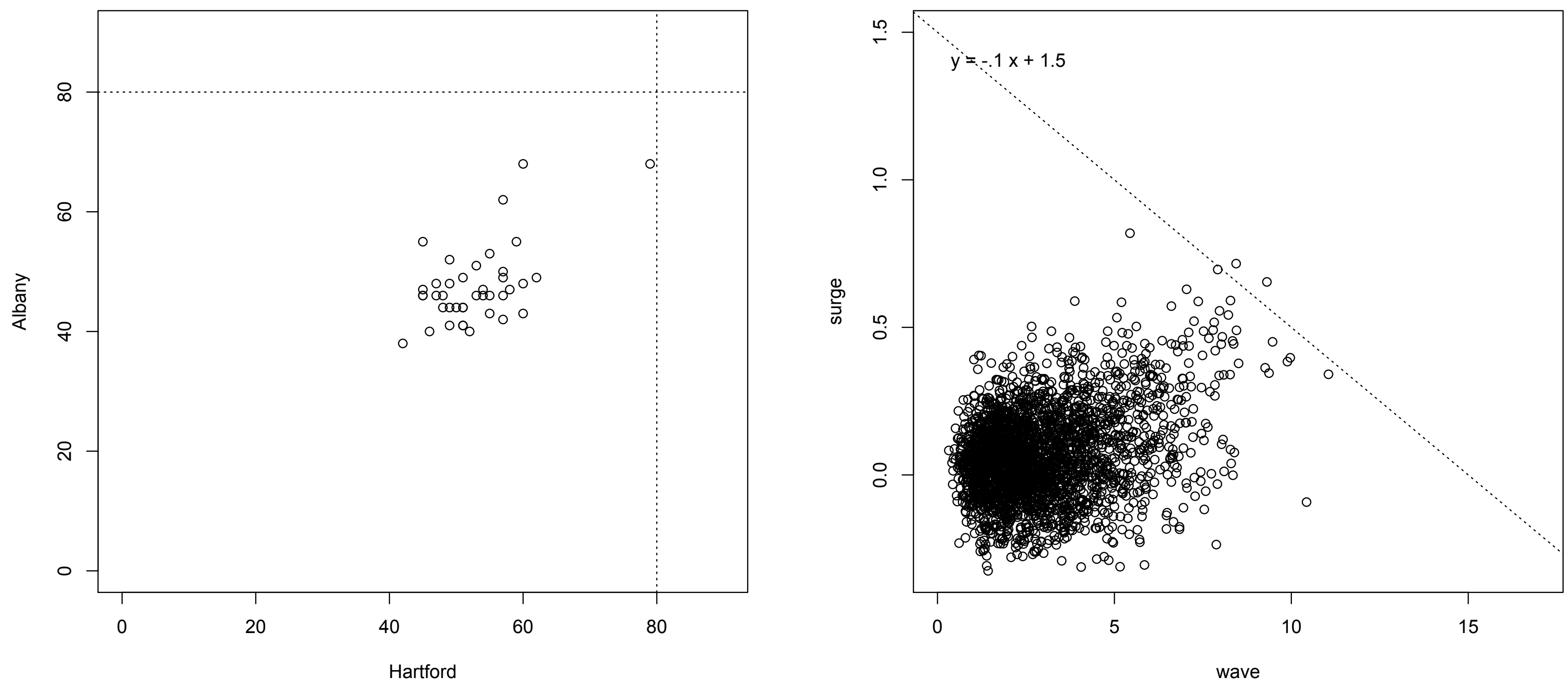
Part IV: Introduction to Bivariate Extremes

1. Preliminaries
   (a) What is the goal of a MV analysis?
   (b) What is meant by tail dependence?
   (c) What is a multivariate extreme?
   (d) Separating marginal and dependence effects.
2. A Probabilistic Framework: Regular Variation
   (a) Definitions and polar coordinate transformation.
   (b) Point process representation and threshold exceedances.
   (c) Properties of the angular measure.
   (d) MV max-stable distributions.
3. Statistics
   (a) Block maximum analysis.
   (b) Threshold exceedance analyses.
4. Dependence Summary Metrics
5. Asymptotic Independence

Foreward: Goal of a MV Extreme Analysis

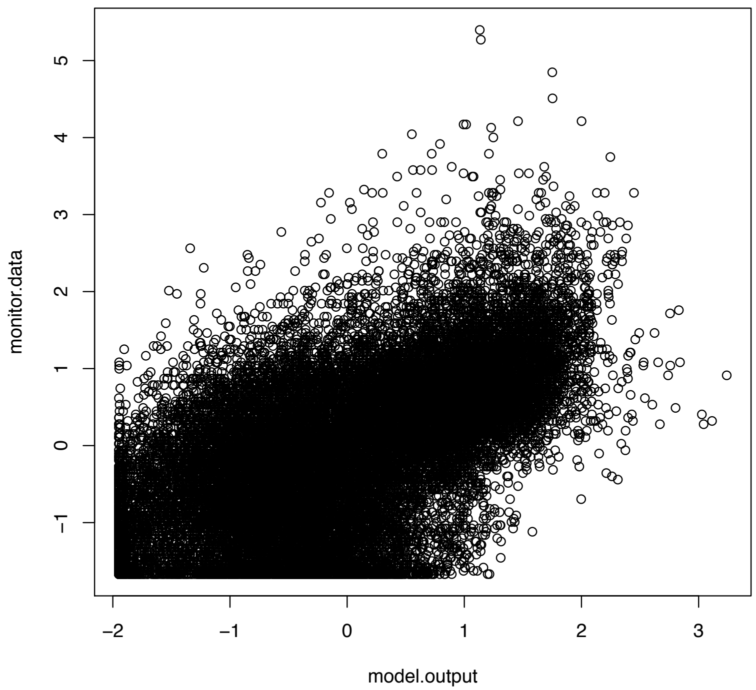
Goal: often to assess probability of falling in a risk region. Sometimes requires extrapolation.

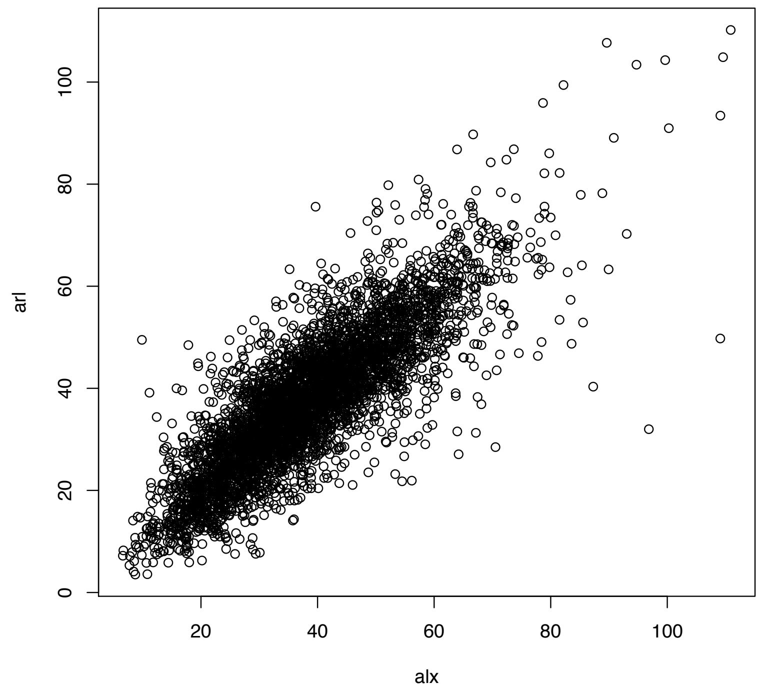
Keep in mind: A basic tenet of an extreme value analysis is to only use data considered to be extreme.



Left: Annual max wind speeds at Hartford and Albany (Coles, 2001).
Right: Wave height and storm surge data (Coles, 2001).

Tail Dependence



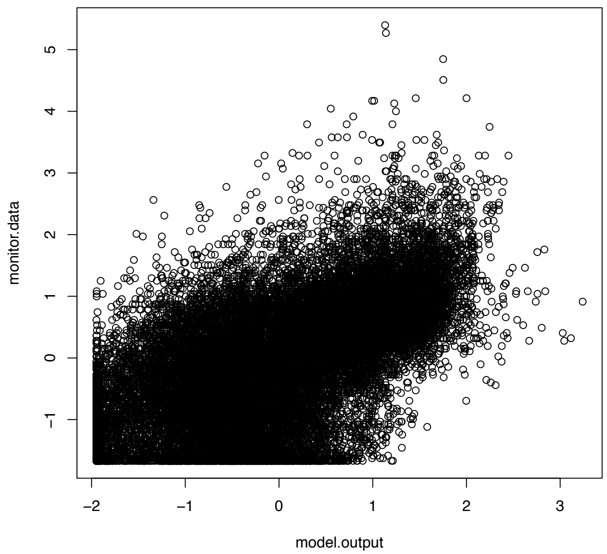


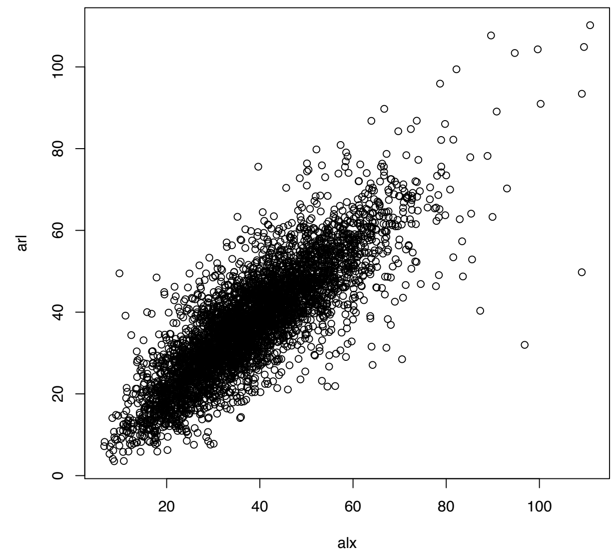
A central aim of multivariate extremes is trying to find an appropriate structure to describe *tail dependence*.

To assess probability of falling in risk region, we need to know how points in the tail behave jointly.

NOT Tail Dependence: Correlation

$$\rho = \frac{E[(X - \mu_x)(Y - \mu_Y)]}{\sqrt{E[(X - \mu_x)^2]E[(Y - \mu_y)^2]}}$$





$\hat{\rho} = 0.59$ $\hat{\rho} = 0.83$

Correlation measures “spread from center”, does not focus on extremes.

A Start: Asymptotic Dependence/Independence

A random vector (X, Y) *with common marginals* is termed asymptotically independent if

$$\lim_{u \to x^+} P(X > u \mid Y > u) = 0.$$

Or if X has cdf F_X and Y has cdf F_Y, then

$$\lim_{u \to 1} P(F_X(X) > u \mid F_Y(Y) > u) = 0.$$

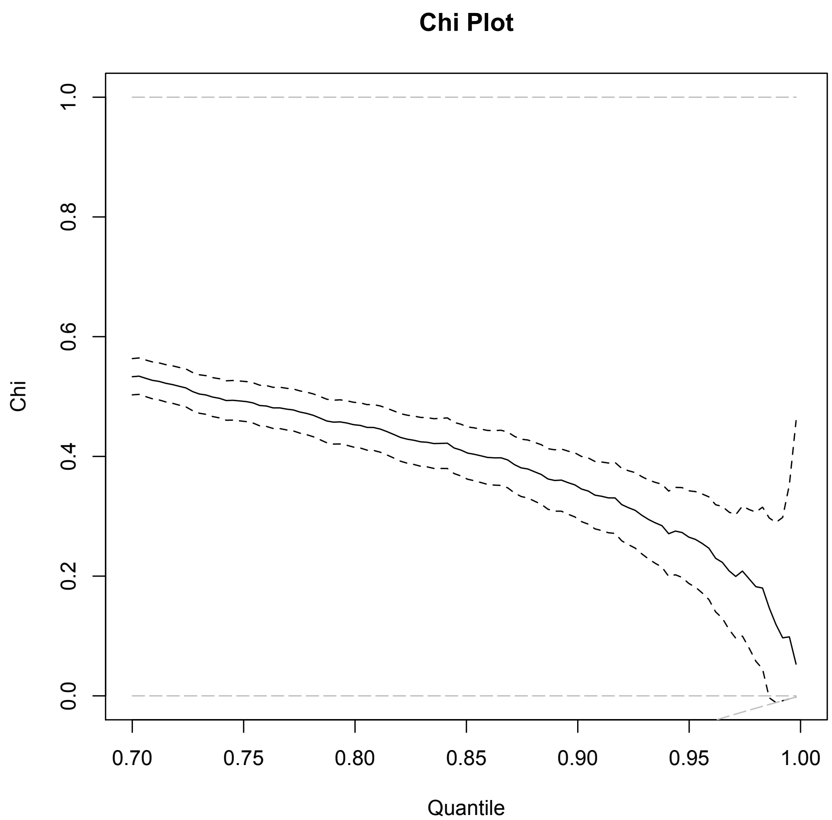
To talk about tail dependence, we need to know something about what it means to be in the tail of each component:

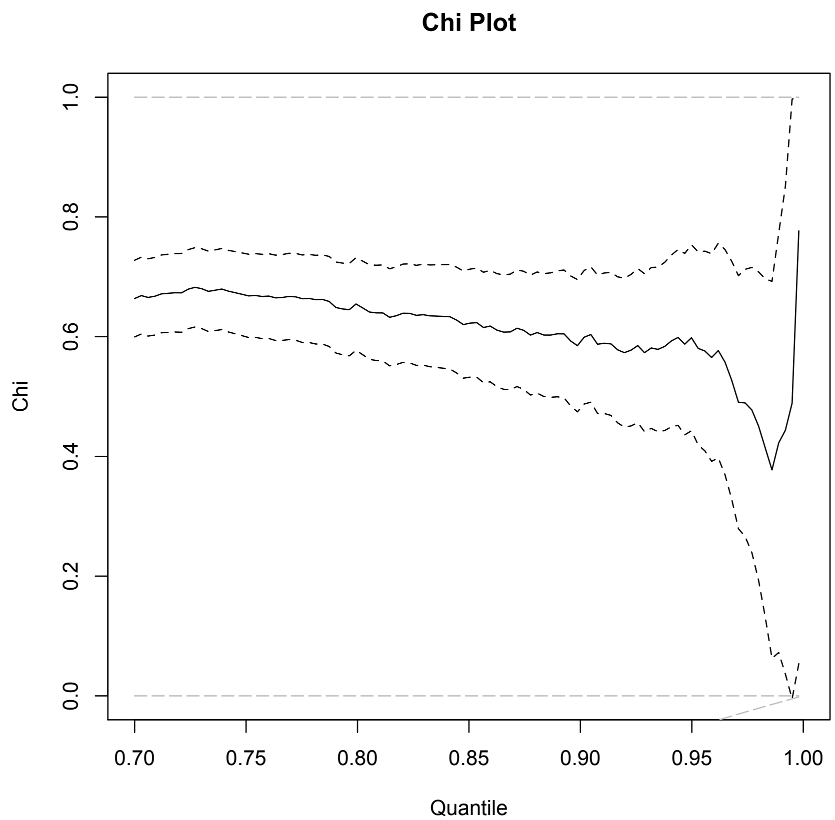
- have a common marginal,
- or account for different marginals.

Asymptotic dependence/independence is a way to *begin* to talk about tail dependence.

Tail Dependence of Examples

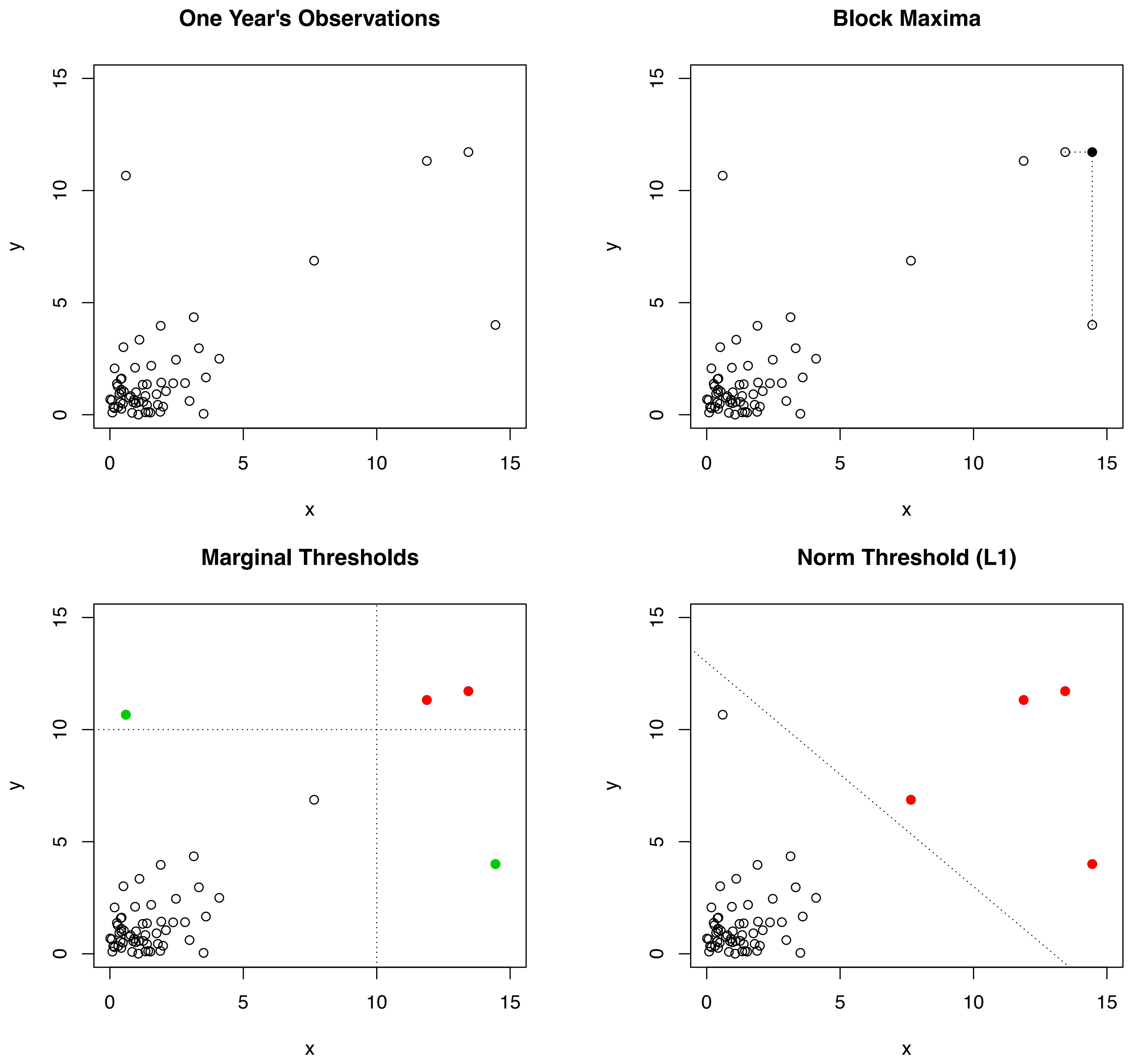
χ is an empirical measure of asymptotic dependence (Coles et al., 1999).





But we will need more than just a summary measure of dependence to *model* the tail–need a probabilistic framework.

What is a Multivariate Extreme?



Multivariate Extremes and Marginal Distributions

In multivariate extremes, dependence is modeled/described after marginal effects have been accounted for.

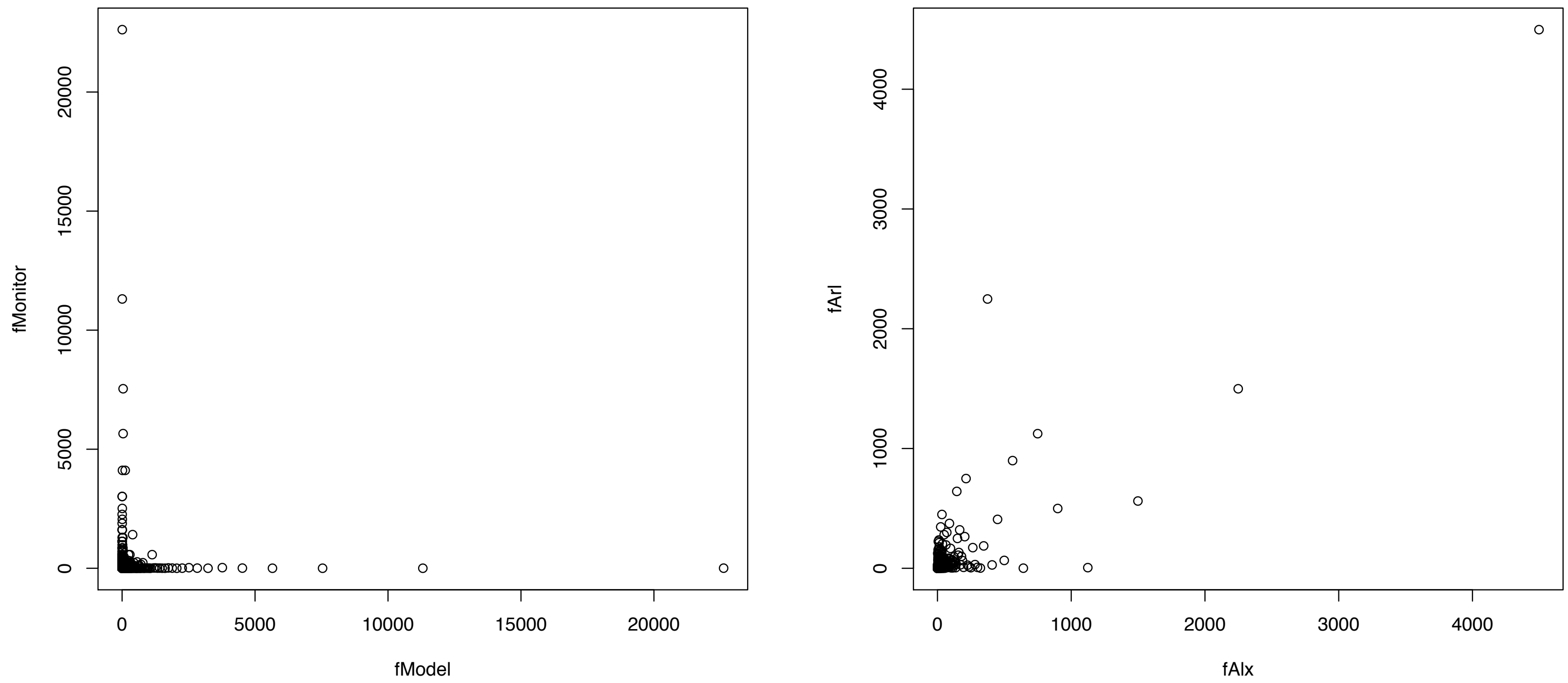
Theory: MV distributions used in extremes are described by first assuming a common marginal distribution, often unit Fréchet ($P(Z \leq z) = \exp(-z^{-1})$). Also (in theory) the marginal distribution *doesn't really matter* when describing dependence because of "domain of attraction" results (see Resnick (1987)).

Practice: In practice, the marginal distributions *do* matter. To apply MV extremal distributions, one must estimate the marginal, and then transform to have common marginals.

Estimation: One can do the two-step process suggested above, or in certain instances, both the marginal distributions and dependence structure can be estimated all-at-once.

Sounds copula-like, but with different marginals and models.

Marginal-transformed Example Data



We will need to look at these heavy-tailed scatterplots differently than we are used to looking at scatterplots.

Example: Bivariate Logistic

The bivariate logistic distribution is regularly varying with tail index 1.

$$F(z_1, z_2) = \exp\left[-\left(z_1^{-1/\beta} + z_2^{-1/\beta}\right)^{\beta}\right]$$

$\beta \in (0, 1]$ controls the amount of dependence, low β implies strong dependence (more on this later).

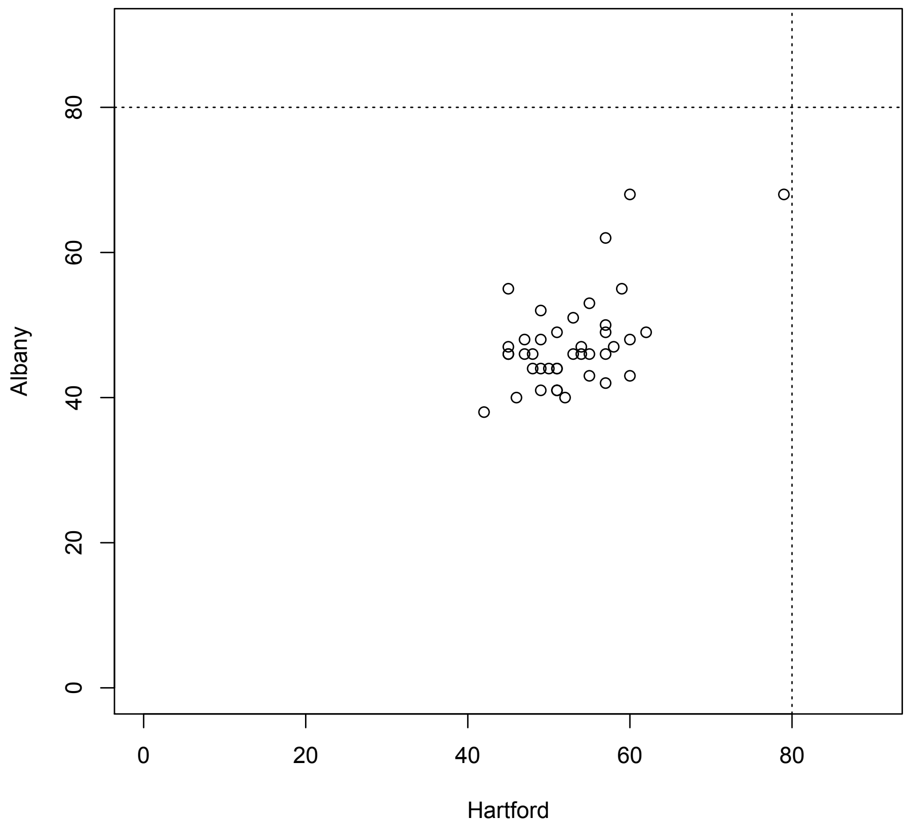
R Demo

Statistics: Fitting a MVEVD

Logistic Model

$$G(z_1, z_2) = \exp\left[-\left(z_1^{-1/\beta} + z_2^{-1/\beta}\right)^{\beta}\right]$$

Annual max wind speed at Hartford and Albany. *R Demo*



Fitted Logistic Model to Wind Data

$\widehat{\mu}_1$	$\widehat{\sigma}_1$	$\widehat{\xi}_1$	$\widehat{\mu}_2$	$\widehat{\sigma}_2$	$\widehat{\xi}_2$	$\widehat{\beta}$
49.97	5.03	0.01	44.58	4.34	0.8	0.71
(0.87)	(0.64)	(0.09)	(0.77)	(0.57)	(0.11)	(0.10)

Note: estimation of angular measure has been done "behind the scenes". Encapsulated in estimate $\widehat{\beta}$.

Estimation of Risk

$$P(M_1 > 80 \text{ or } M_2 > 80) \stackrel{\text{est}}{=} 0.0042$$
$$P(M_1 > 80 \text{ and } M_2 > 80) \stackrel{\text{est}}{=} 0.00086$$
$$P(M_1 > 80)P(M_2 > 80) \stackrel{\text{est}}{=} 0.000006$$

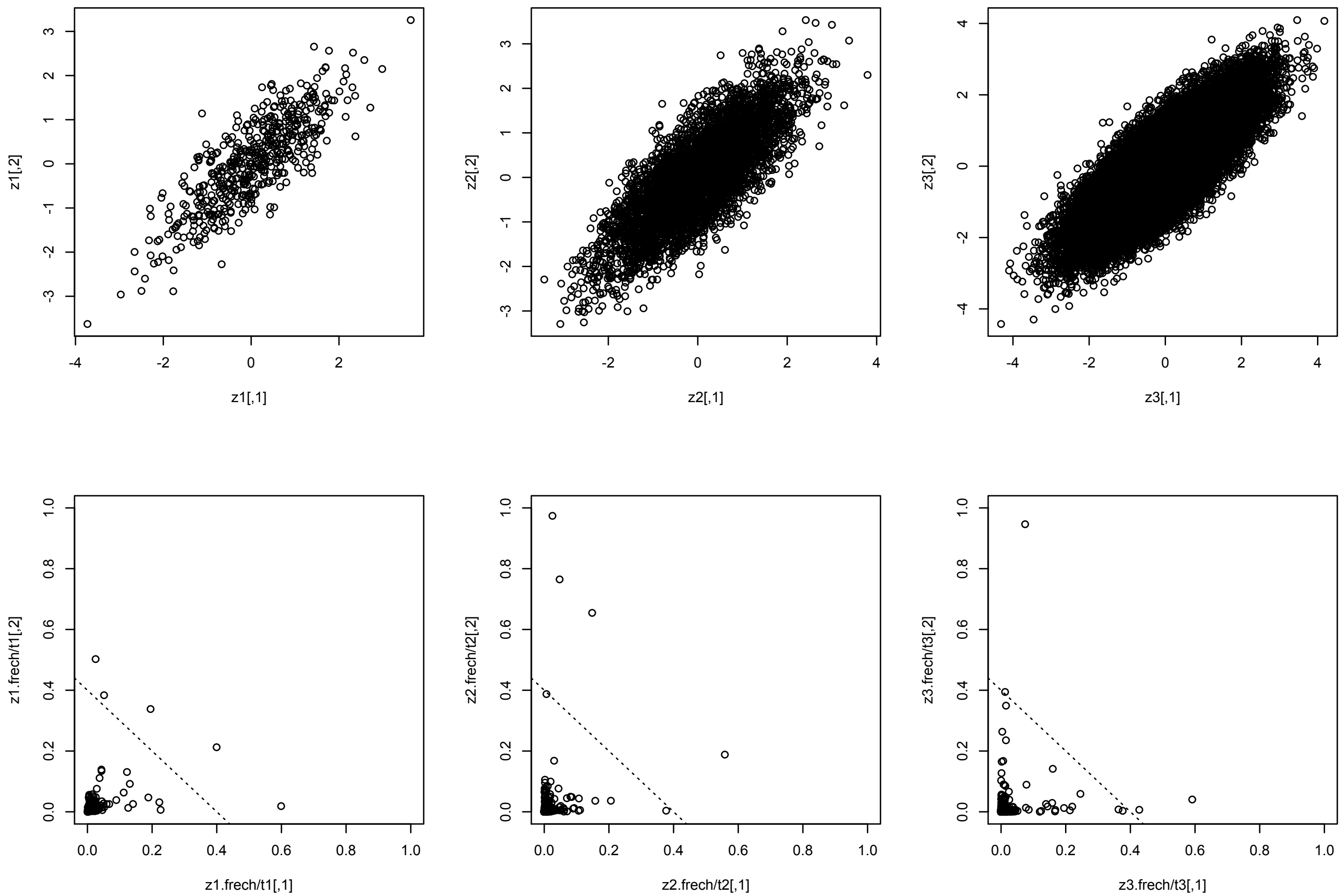
There is dependence in this data. Note the difference between the "joint" and "independent" estimates.

Asymptotic Independence

Bivariate Normal to Fréchet Margins Example

R Demo Here

Bivariate Normal to Fréchet Margins Example



Asymptotic Independence:

(X, Y) (with common marginals) are said to be asymptotically independent if $\lim_{x \to x_+} \mathbb{P}(Y > x | X > x) = 0$.

Take-away messages: Part IV

- Definition of a multivariate extreme is not obvious.
- Tail dependence is different than what we usually think of as dependence.
- Current methodology separately handles marginal effects and dependence.
- Regular variation provides a mathematical framework–leads to a polar decomposition.
- In regular variation framework, tail dependence is completely described by the angular measure.
- Tail dependence *not* summarized with correlations, we looked at the extremal coefficient.
- Methodologies exist for both block maxima and threshold exceedance approaches.
- Regular variation (and classical MV EVT) requires extension to describe the dependence in the asymptotic independence case.

Take-away messages (overall)

- An extreme value analysis uses only data considered to be extreme. Inference about the tail can be contaminated by data that is not extreme.
- Distributions for tail modeling are justified by asymptotic results from probability theory. These give us rationale to extrapolate beyond the range of the data.
- One is always data poor when doing extreme value analyses. Large uncertainties are intrinsic to the problem.
- Tail dependence is described very differently than dependence in the central part of the distribution.

Things not Addressed

- Rates of convergence to the limiting distributions.
- Multivariate models for $d > 2$.
- Spatial extremes.
- Bayesian inference.

References

Beirlant, J., Caeiro, F., and Gomes, M. (2012). An overview and open research topics in statistics of univariate extremes. *REVSTAT–Statistical Journal*, 10(1):1–31.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., Waal, D. D., and Ferro, C. (2004). *Statistics of Extremes: Theory and Applications*. Wiley, New York.

Chavez-Demoulin, V. and Davison, A. (2005). Generalized additive models for sample extremes. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 54(1):207–222.

Coles, S., Heffernan, J., and Tawn, J. (1999). Dependence measures for extreme value analysis. *Extremes*, 2:339–365.

Coles, S. and Tawn, J. (1991). Modeling multivariate extreme events. *Journal of the Royal Statistical Society, Series B*, 53:377–92.

Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. Springer-Verlag London Ltd., London.

Cooley, D. (2009). Extreme value analysis and the study of climate change, commentary on Wigley 1988. *Climatic Change*, 97:77–83.

Cooley, D. (2012). Return periods under climate change. In AghaKouchak, A., Easterling, D., Hsu, K., Schubert, S., and Sorooshian, S., editors, *Extremes in a Changing Climate: Detection, Analysis and Uncertainty*, Water Science and Technology Series, Vol 65, pages 97–114. Springer, Netherlands, Dordrecht.

Cooley, D., Naveau, P., and Poncet, P. (2006). Variograms for spatial max-stable random fields. In Bertail, P., Doukhan, P., and Soulier, P., editors, *Dependence in Probability and Statistics*, Springer Lecture Notes in Statistics. Springer, New York.

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.

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

De Haan, L. and Zhou, C. (2011). Extreme residual dependence for random vectors and processes. *Advances in Applied Probability*, 43(1):217–242.

Fawcett, L. and Walshaw, D. (2007). Improved estimation for temporally clustered extremes. *Environmetrics*, 18(2):173–188.

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society, Series B*, 66:497–546.

Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, pages 1163–1174.

Leadbetter, M., Lindgren, G., and Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. Springer Verlag, New York.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83:169–187.

Maulik, K. and Resnick, S. (2004). Characterizations and examples of hidden regular variation. *Extremes*, 7(1):31–67.

Ramos, A. and Ledford, A. (2009). A new class of models for bivariate joint tails. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(1):219–241.

Reich, B., Cooley, D., Foley, K., Napelenok, S., and Shaby, B. (2012). Extreme value analysis for evaluating ozone control strategies. *Submitted*, XX:XX–XX.

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

Resnick, S. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5(4):303–336.

Resnick, S. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer Series in Operations Research and Financial Engineering. Springer, New York.

Sang, H. and Gelfand, A. E. (2009). Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics*, 16:407–426.

Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold es-timation and uncertainty quantification. *REVSTAT–Statistical Journal*, 10(1):33–60.

Schlather, M. and Tawn, J. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika*, 90:139–156.

Smith, R. (1990). Max-stable processes and spatial extremes. Technical report, University of North Carolina, Chapel Hill.

Weller, G. and Cooley, D. (2013). A sum decomposition for hidden regular variation in joint tail modeling with likelihood inference via the mcem algorithm. *Submitted*.