

HIERARCHICAL MODELS IN EXTREME VALUE THEORY

Richard L. Smith

Department of Statistics and Operations Research,
University of North Carolina, Chapel Hill

Intro Lecture for STOR 834

Based on talks previously presented at NCSU (February 2015),
UNC (August 2018) and a few other places

OUTLINE

- I. An Example of Hierarchical Models Applied to Insurance Extremes
- II. Attribution of Climate Extremes
- III. Joint Distributions of Climate Extremes

I. AN EXAMPLE OF HIERARCHICAL MODELS APPLIED TO INSURANCE EXTREMES

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

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

See also:

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

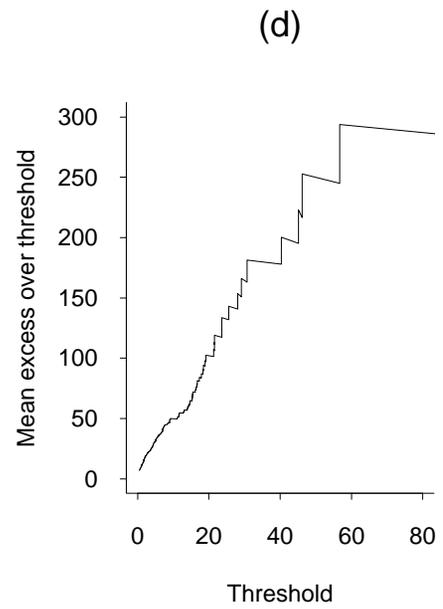
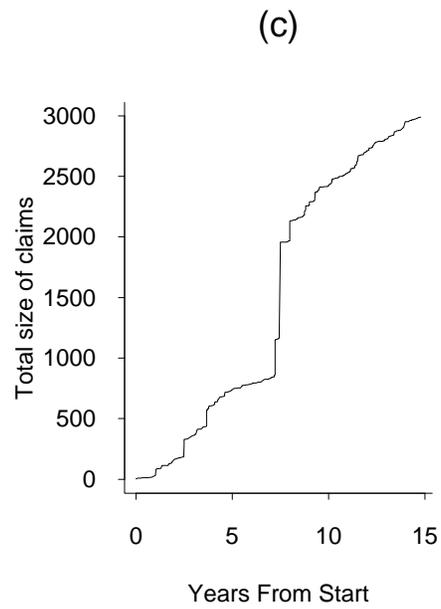
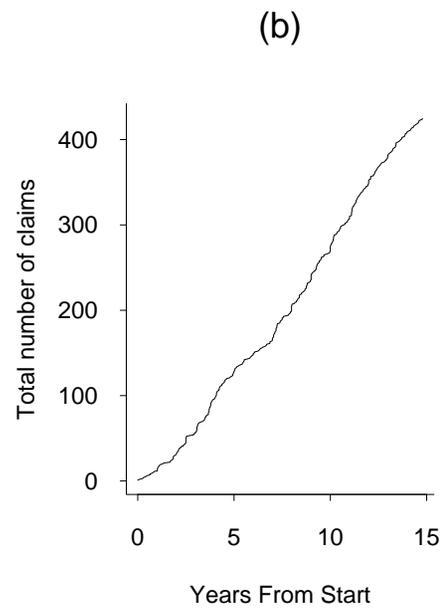
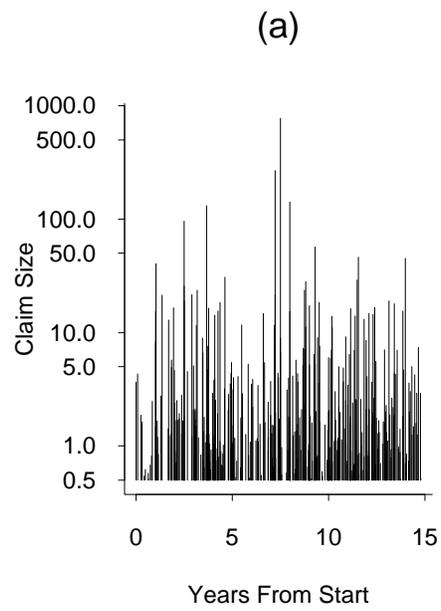
<http://www.stat.unc.edu/postscript/rs/semstatrls.pdf>

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

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

Total of all 393 claims: 2989.6

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



Some plots of the insurance data.

Some problems:

1. What is the distribution of very large claims?
2. Is there any evidence of a change of the distribution over time?
3. What is the influence of the different types of claim?
4. How should one characterize the risk to the company? More precisely, what probability distribution can one put on the amount of money that the company will have to pay out in settlement of large insurance claims over a future time period of, say, three years?

Introduction to Univariate Extreme Value Theory

EXTREME VALUE DISTRIBUTIONS

X_1, X_2, \dots , i.i.d., $F(x) = \Pr\{X_i \leq x\}$, $M_n = \max(X_1, \dots, X_n)$,
 $\Pr\{M_n \leq x\} = F(x)^n$.

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

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

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

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

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. $\xi \rightarrow 0$ 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 $u \rightarrow \omega_F = \sup\{x : F(x) < 1\}$, often find a limit

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

where G is *generalized Pareto distribution* (GPD)

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

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

The Generalized Pareto Distribution

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

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

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

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

i.e. exponential distribution with mean σ ,

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

POISSON-GPD MODEL FOR EXCEEDANCES

1. The number, N , of exceedances of the level u in any one year has a Poisson distribution with mean λ ,
2. Conditionally on $N \geq 1$, the excess values Y_1, \dots, Y_N are IID from the GPD.

Relation to GEV for annual maxima:

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

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

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

ALTERNATIVE PROBABILITY MODELS

1. The r largest order statistics model

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

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

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

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

2. Point process approach (Smith 1989)

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

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

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

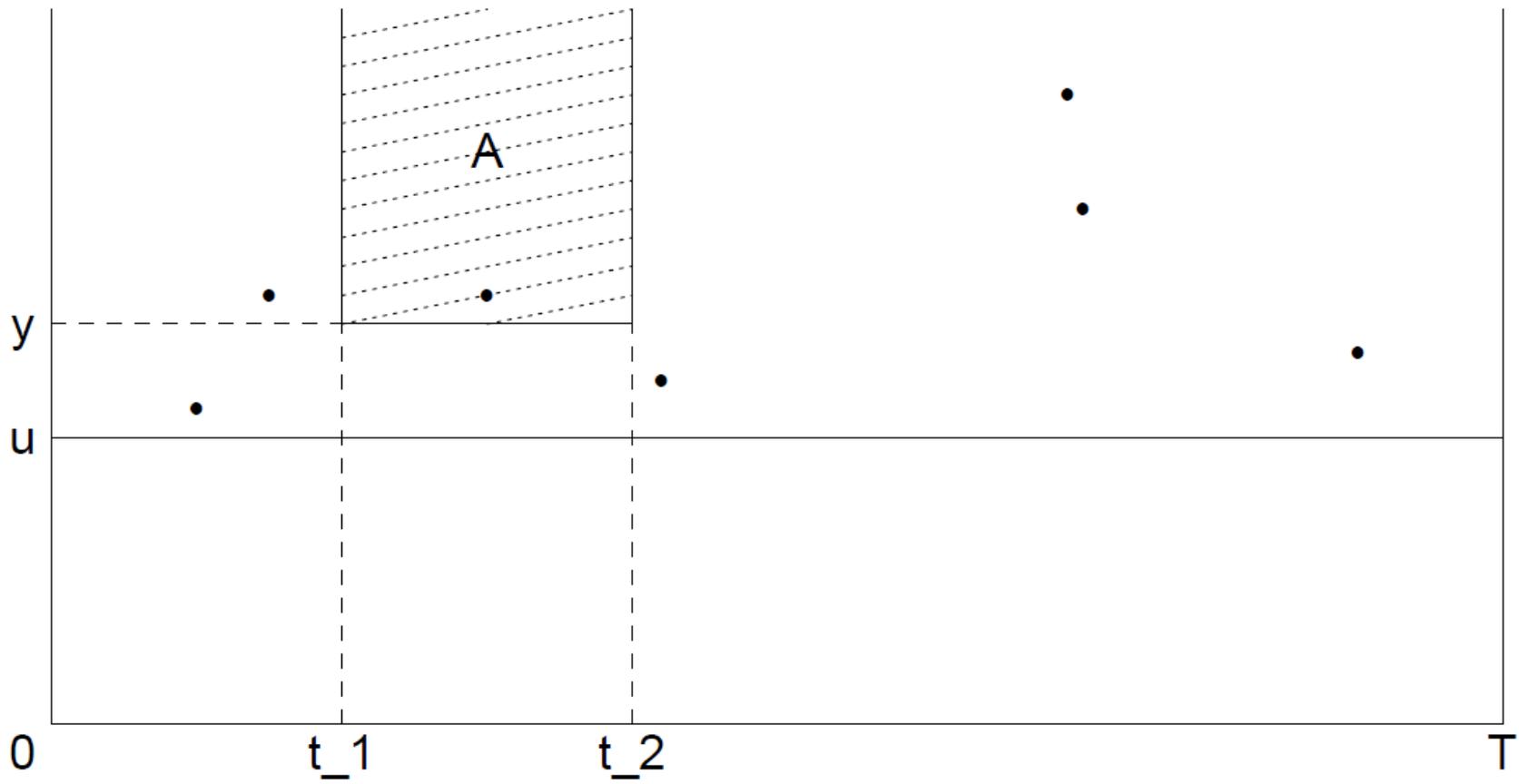


Illustration of point process model.

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

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

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

ESTIMATION

GEV log likelihood:

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

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

Poisson-GPD model:

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

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

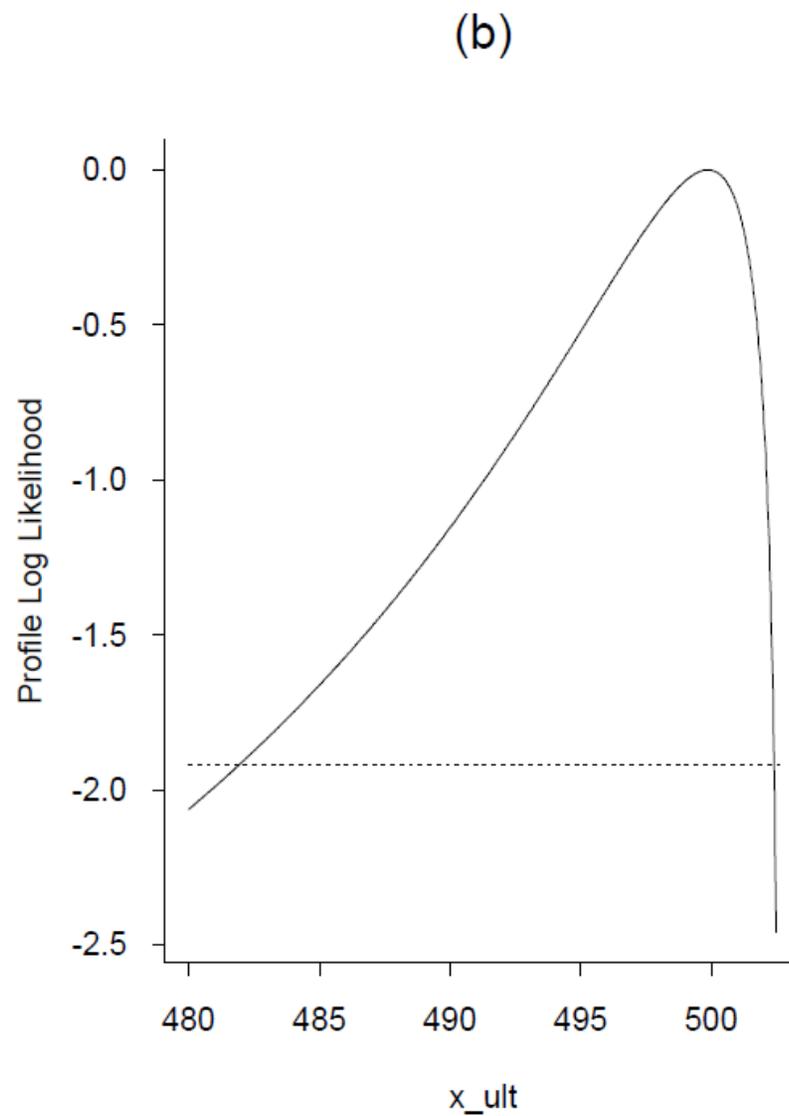
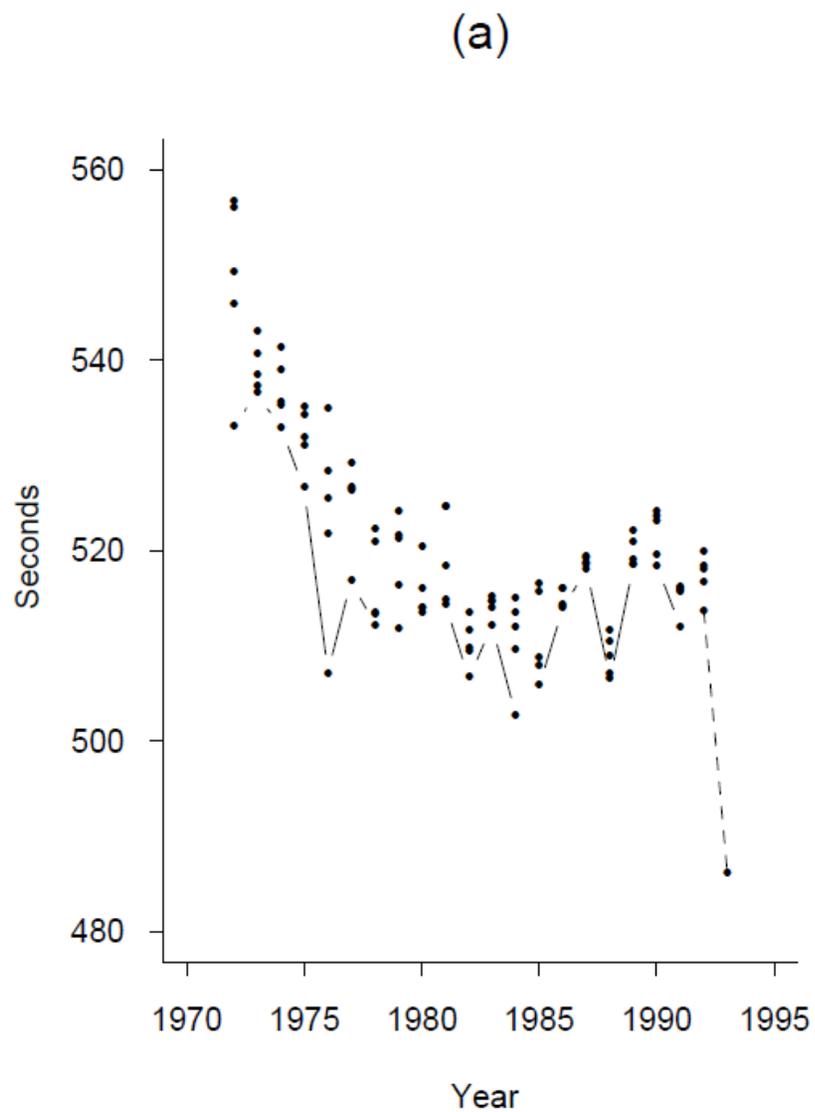
Usual asymptotics valid if $\xi > -\frac{1}{2}$ (Smith 1985)

Bayesian approaches

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

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

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



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

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

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

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

Alternative Bayesian calculation:

(Smith 1997)

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

The answer is approximately 0.0004.

Insurance Extremes Dataset

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

GPD fits to various thresholds:

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

Point process approach:

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

Standard errors are in parentheses

Predictive Distributions of Future Losses

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

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

Traditional frequentist approach:

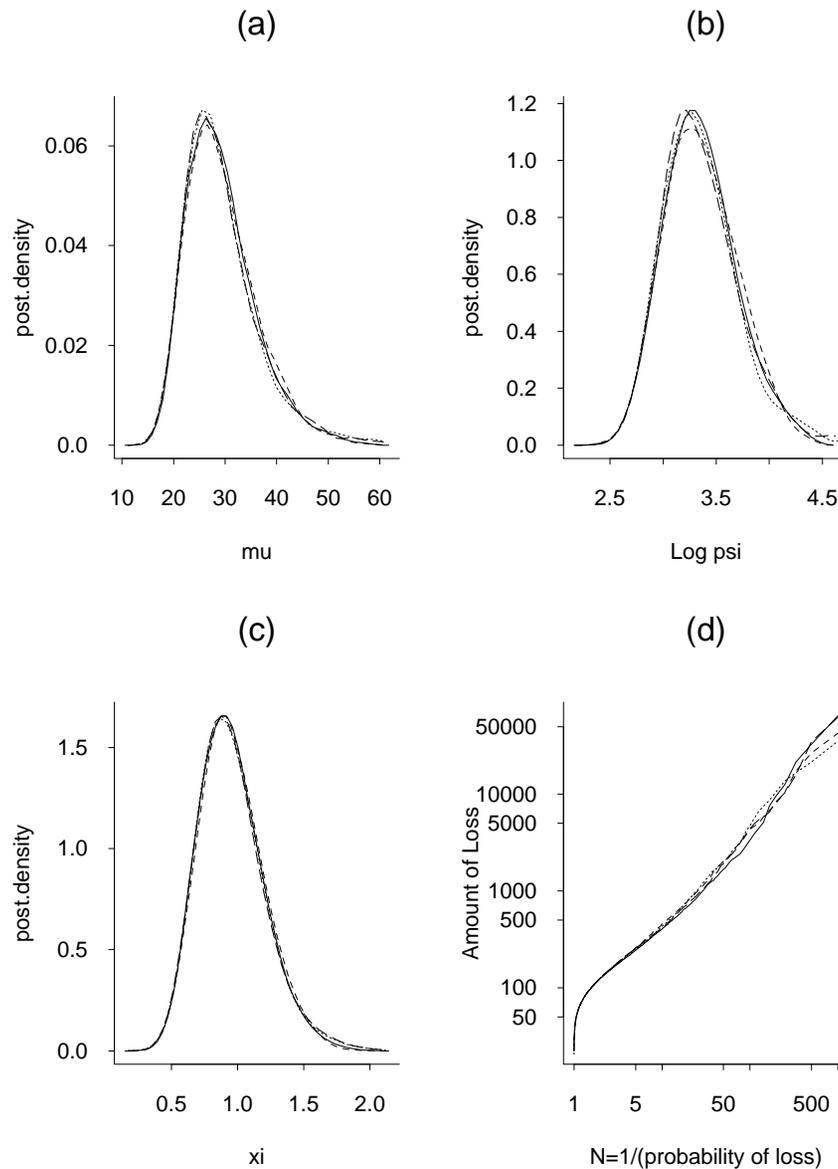
$$\hat{G}(y) = G(y; \hat{\mu}, \hat{\psi}, \hat{\xi})$$

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

Bayesian:

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

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



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

Hierarchical models for claim type and year effects

Further features of the data:

1. When separate GPDs are fitted to each of the 6 main types, there are clear differences among the parameters.
2. The rate of high-threshold crossings does not appear uniform, but peaks around years 10–12.

A Hierarchical Model:

Level I. Parameters m_μ , m_ψ , m_ξ , s_μ^2 , s_ψ^2 , s_ξ^2 are generated from a prior distribution.

Level II. Conditional on the parameters in Level I, parameters μ_1, \dots, μ_J (where J is the number of types) are independently drawn from $N(m_\mu, s_\mu^2)$, the normal distribution with mean m_μ , variance s_μ^2 . Similarly, $\log \psi_1, \dots, \log \psi_J$ are drawn independently from $N(m_\psi, s_\psi^2)$, ξ_1, \dots, ξ_J are drawn independently from $N(m_\xi, s_\xi^2)$.

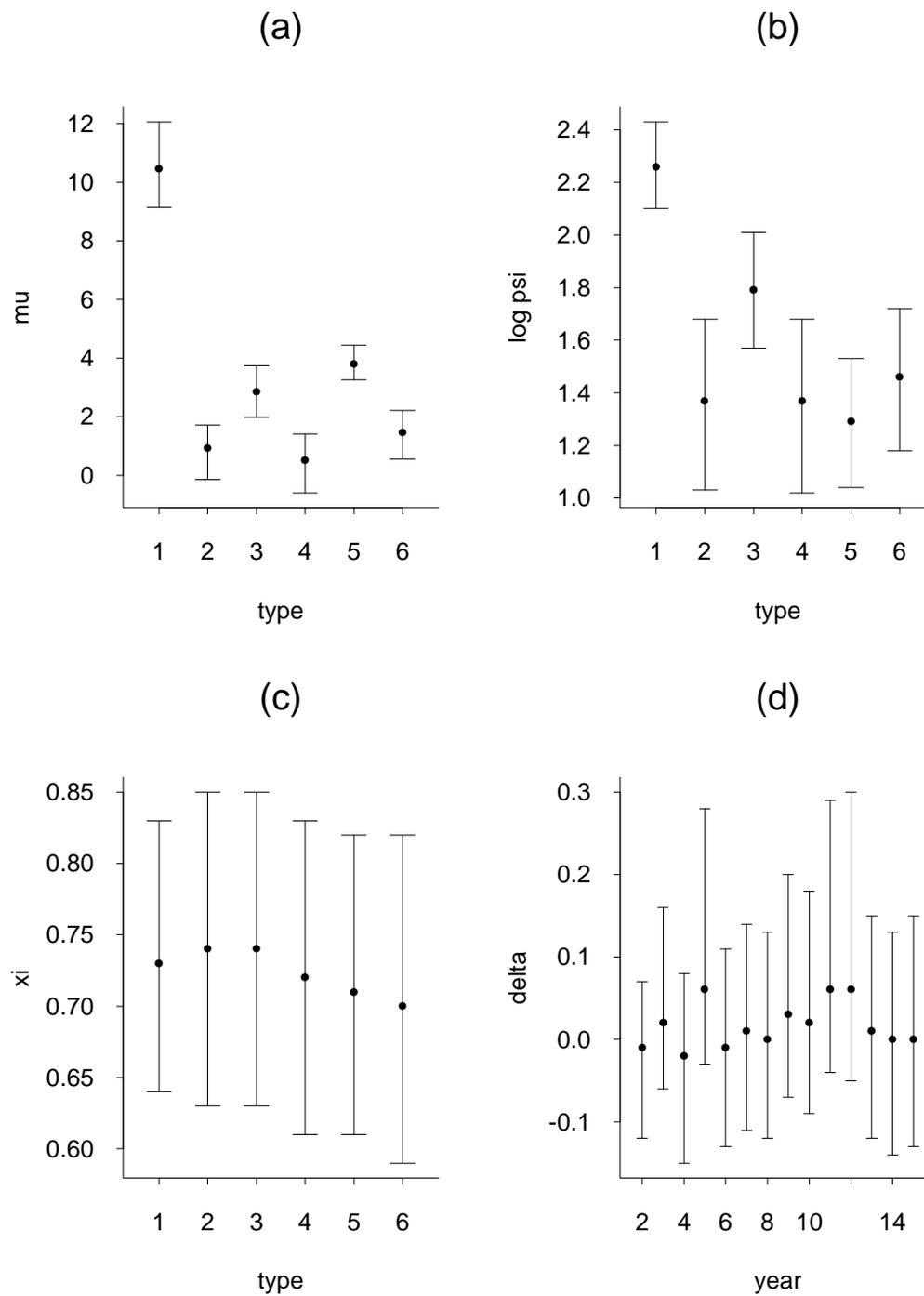
Level III. Conditional on Level II, for each $j \in \{1, \dots, J\}$, the point process of exceedances of type j is generated from the Poisson process with parameters μ_j , ψ_j , ξ_j .

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

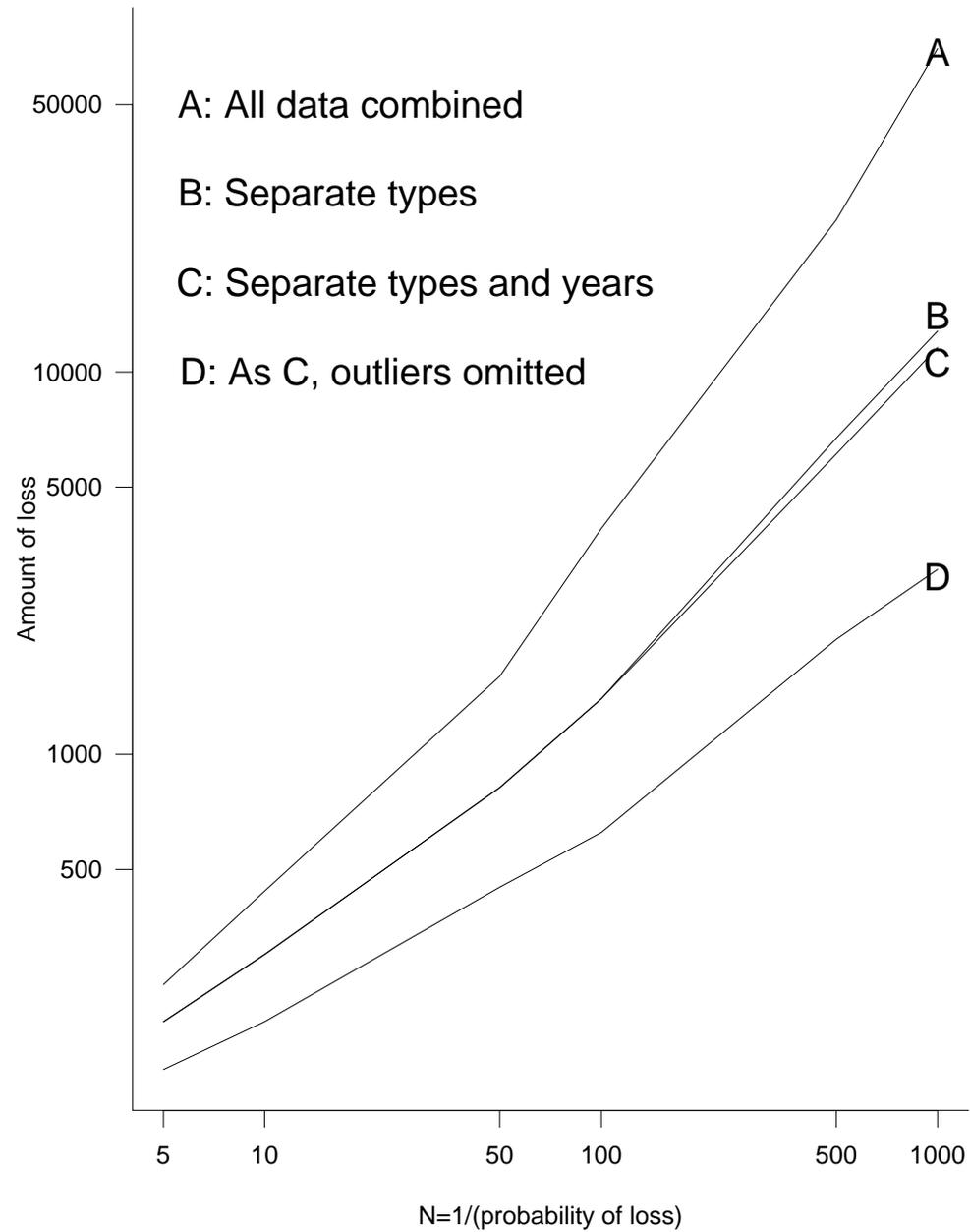
$$\delta_k = \rho\delta_{k-1} + \eta_k, \quad \eta_k \sim N(0, s_\eta^2)$$

with a vague prior on (ρ, s_η^2) .

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



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



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

Daniel Cooley, Brett D. Hunter, Richard L. Smith

***Univariate and Multivariate
Extremes for the Environmental
Sciences***

<https://rls.sites.oasis.unc.edu/postscript/rs/8-Cooley-Extremes.pdf>

Example from Cooley et al. (2019)

- Climate model output (CESM1) — “initial condition ensemble” with 30 members
- Use historical runs from 1920–2005 though they are also extended to 2100 under RCP 4.5 and RCP 8.5
- Problem: Does extreme precipitation show a time trend
- Notation: $M_b^{(j)}(s)$ annual max daily precipitation for year b at location s in ensemble member j
- GEV model $G(y) = \exp \left\{ - \left(1 + \xi \cdot \frac{y - \eta}{\tau} \right)_+^{-1/\xi} \right\}$, density g

First model

- $M_b^{(j)}(\mathbf{s}) \sim GEV(\eta_b^{(j)}(\mathbf{s}), \tau^{(j)}(\mathbf{s}), \xi^{(j)}(\mathbf{s}))$ where $\eta_b^{(j)}(\mathbf{s}) = \beta_0^{(j)}(\mathbf{s}) + \beta_1(\mathbf{s})^{(j)}(b - 1919)$
- Fit separate model for each j and \mathbf{s}
- Fig. 1 shows $\beta_1^{(j)}(\mathbf{s})$ and $\xi^{(j)}(\mathbf{s})$ for each j for one grid cell \mathbf{s} (Fort Collins, CO) and each ensemble member j
- Overall: wide variability in both parameters

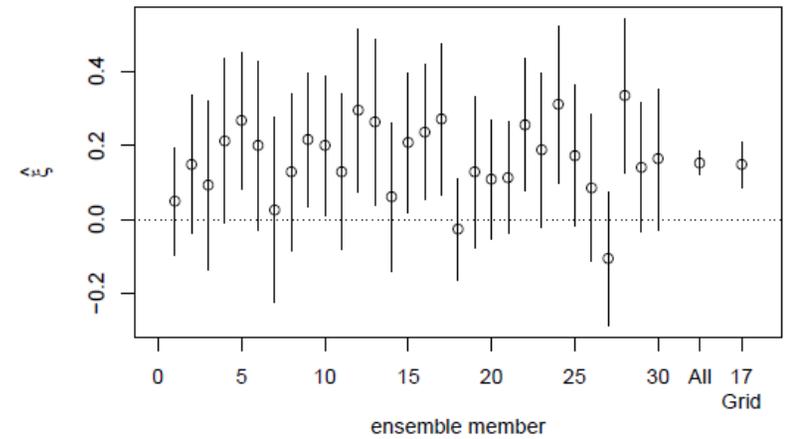
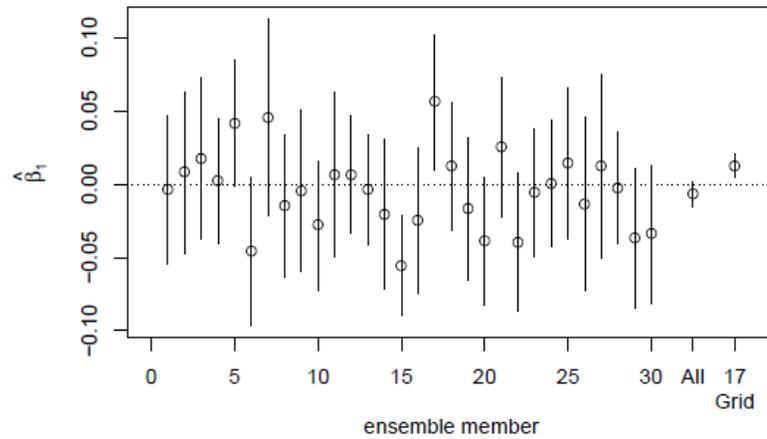


FIGURE 1: Estimates for β_1 and ξ for the Fort Collins grid cell from individual ensemble members. “All” on horizontal axes indicates when all ensemble members were used. “17 Grid” indicates only data from ensemble member 17 was used and the method to borrow strength across location in Section 0.2.5.2 was implemented.

Alternative Viewpoints

- Look just at ensemble member 17 (Fig. 2, top left) with fitted trend lines), or
- Combine all ensemble members together (Fig. 2, top right), or
- Look at just ensemble member 17 but with neighboring grid cells (Fig. 2, bottom)
- Most climate model outputs do not contain multiple ensembles, so second approach may not be practical for all applications.

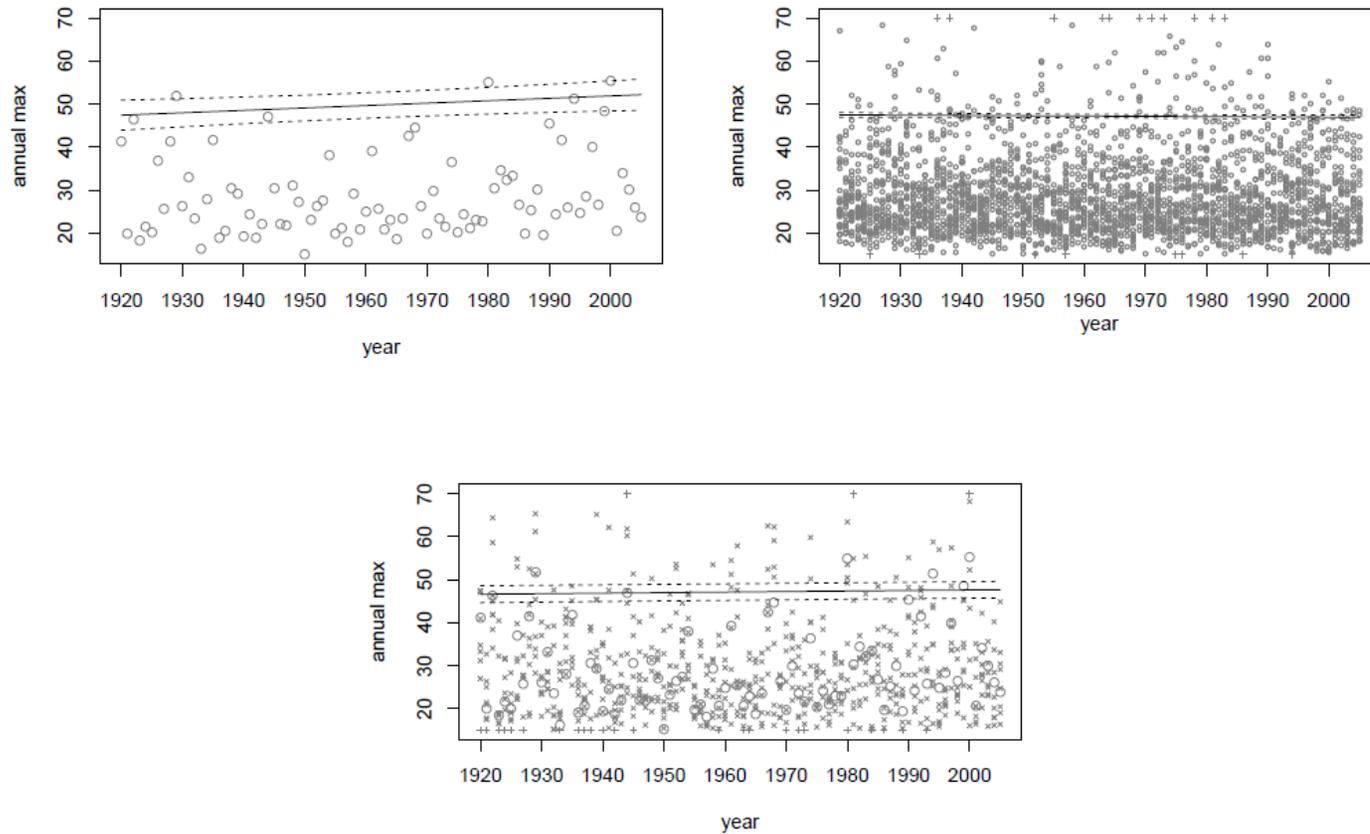


FIGURE 2: Top left: Circles denote annual maximum daily precipitation (mm) for Fort Collins grid cell from ensemble member 17. Top right: Same but for all ensembles. Bottom center shows ensemble 17's annual maximum precipitation for Fort Collins grid cell with circles and annual maxima from eight surrounding grid cells with X's. Plotted in each figure is the point estimate for the .95 quantile (equivalently AEP(.05) level) along with 95% confidence intervals. Point estimates for the top left, top right, and bottom center respectively use only ensemble 17's data for the Fort Collins grid cell, all ensembles' data for the Fort Collins grid cell, and the borrowing strength estimate described in Section 0.2.5.2. In the right and bottom plots, values greater than 70 mm or less than 15 mm fall outside the plotting window and have been denoted by crosses.

Borrowing strength across locations

- One ensemble member ($j = 17$)
- Grid cells s_1, \dots, s_9
- Assume $\beta_0(s)$, $\tau(s)$ (different for $s = s_1, \dots, s_9$) but β_1 and ξ same for all locations
- Maximize log likelihood $\ell(\beta_0, \beta_1, \tau, \xi)$
 $= \prod_{i=1}^9 \prod_{b=1920}^{2005} g(M_b(s_i); \eta_b(s_i) = \beta_0(s_i) + \beta_1(b-1919), \tau(s_i), \xi)$
- In this model, the 95% confidence interval for β_1 includes 0 (no significant trend)

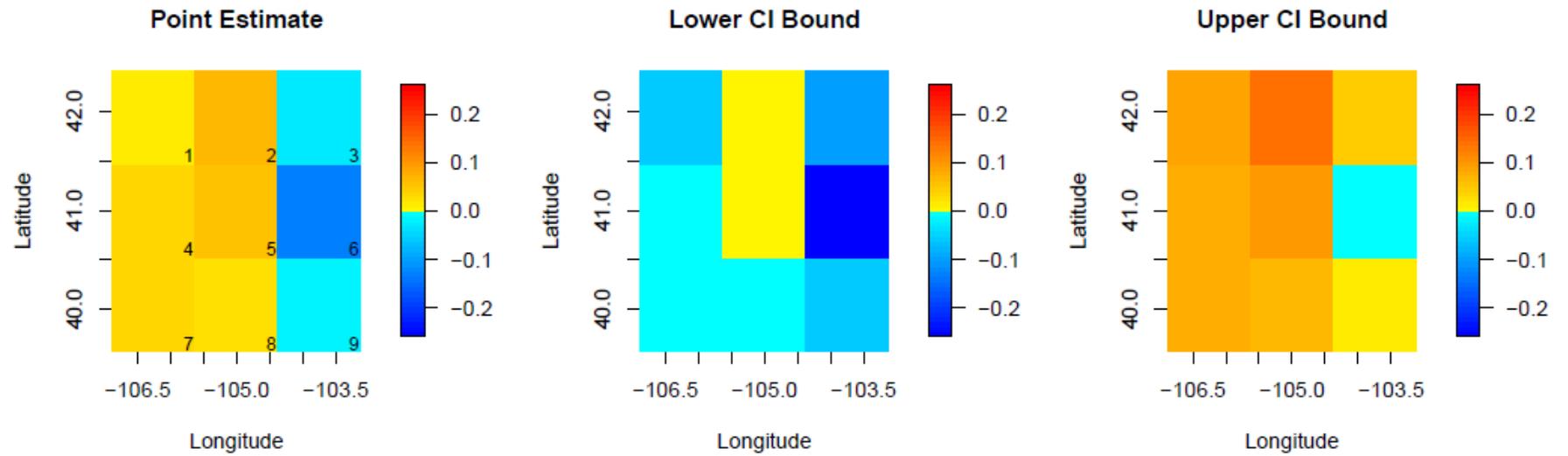


FIGURE 3: Point estimates (left), lower 95% CI bound (center), and upper 95% CI bound (right) for $\beta_1(s)$ (mm/year). Yellow-to-red colors indicate positive values while turquoise-to-blue indicate negative values.

A more general model

- We could extend this to a model for all locations s simultaneously
- $\beta_0(s), \beta_1(s), \tau(s), \xi(s)$ form a 4-dimensional Gaussian spatial process
- A hierarchical model would allow the parameters of the full spatial process to be estimated simultaneously with the individual parameters $\beta_0(s), \beta_1(s), \tau(s), \xi(s)$
- Applications include looking for trends in extremes, estimating extreme quantiles, etc.
- If we did this simultaneously for models that do or do not include anthropogenic forcing factors, this could be a basis for “attribution of extremes”

Introduction to Multivariate Extremes

- Simplest model: 2-dimensional (Y_1, Y_2) with marginal CDFs F_1, F_2
- Tail dependence parameter χ (Coles et al, 1999) defined by
$$\chi = \lim_{u \rightarrow 1} P(F_1(Y_1) > u \mid F_2(Y_2) > u)$$
- $\chi = 0$ is *asymptotically independent* case, $\chi > 0$ is *asymptotically dependent*.
- Earliest examples were always for $\chi > 0$ but in recent years it has become recognized that $\chi = 0$ is both harder and more important in practice

Multivariate EVDs and componentwise block maxima

- Assume samples $(Y_{i,1}, Y_{i,2})$, $i = 1, \dots, n$ independent for each i but dependent between the two components

- Componentwise maxima $\mathbf{M}_n = (\max_{i=1}^n Y_{i,1}, \max_{i=1}^n Y_{i,2})$

- Limit laws of form

$$\frac{\mathbf{M}_n - \mathbf{b}_n}{\mathbf{a}_n} \rightarrow G \quad (\text{in distribution})$$

- G is said to be a bivariate extreme value distribution (BVEVD)
- Obvious generalization to more than two dimensions (MVEVD)

Characterization of MVEVDs

- Without loss of generality, we may transform the marginal CDFs to anything we like. The theory is simplified if we assume *unit Fréchet* margins, $F_i(y) = e^{-1/y}$ for $0 < y < \infty$
- In that case, $\mathbf{b}_n = \mathbf{0}$ and $a_n = (n, n)$ (in dimension 2)
- $G(\mathbf{y}) = \exp\{-V(\mathbf{y})\}$ where V is *homogeneous of degree -1* : $V(s\mathbf{y}) = s^{-1}V(\mathbf{y})$ for all $s > 0$.
- Example with $d = 2$, $G(\mathbf{y}) = \exp\left\{-\left(y_1^\beta + y_2^\beta\right)^{-1/\beta}\right\}$ for $y_1 > 0$, $y_2 > 0$, $0 < \beta \leq 1$, known as the *logistic model*
- Simplest case of a large class of parametric BVEVD models — suggests fitting by MLE etc. However, this whole approach is now recognized as too restrictive to cover the full range of multivariate extreme behaviour that we want

Background 1: Probability Integral Transforms

- Suppose $U \sim \text{Unif}(0, 1)$ and F is a CDF. Let $X = F^{-1}(U)$. Then the CDF of X is F .
 - Remark: This does not require F be continuous, so long as you are careful about the definition of F^{-1} in this case.
- Suppose X is a random variable with distribution function F , where F is continuous. Then $F(X)$ has a uniform distribution on $(0, 1)$.
- Suppose $U \sim \text{Unif}(0, 1)$ and $Z = -\frac{1}{\log(U)}$. Then Z has a unit Fréchet distribution (CDF $e^{-1/z}$, $0 < z < \infty$).
- If Y has a GEV distribution with parameters (μ, ψ, ξ) then $Z = \left(1 + \xi \frac{Y - \mu}{\psi}\right)^{1/\xi}$ has a unit Fréchet distribution. In particular, the case $\mu = \psi = \xi = 1$ is unit Fréchet.

Background 2: Max-Stability

- Suppose \mathbf{Z} has unit Fréchet margins, so $\mathbf{a}_n = (n, n)$ and $\mathbf{b}_n = \mathbf{0}$.
- $P^n \left(\frac{\mathbf{Z}}{n} \leq \mathbf{y} \right) \rightarrow e^{-V(\mathbf{y})}$.
- For fixed k as $n \rightarrow \infty$, $P^{nk} \left(\frac{\mathbf{Z}}{nk} \leq \mathbf{y} \right) \rightarrow e^{-V(\mathbf{y})}$.
- Also write as $\left\{ P^n \left(\frac{\mathbf{Z}}{n} \leq k\mathbf{y} \right) \right\}^k \rightarrow e^{-kV(k\mathbf{y})}$.
- So $V(k\mathbf{y}) = k^{-1}V(\mathbf{y})$ for any positive integer k , all $\mathbf{y} > \mathbf{0}$
- Hence $V(s\mathbf{y}) = s^{-1}V(\mathbf{y})$ for any positive rational s , all $\mathbf{y} > \mathbf{0}$
- By rational approximation, this holds for irrational s as well.

Threshold Approaches

- Alternate viewpoint is just to plot all the datapoints (rather than block maxima) — look for evidence of dependence in upper tails
- Unit Fréchet transformation useful to visually what is going on
- Simulated example from logistic model
- Generate $\mathbf{Z}_i = (z_{i,1}, z_{i,2})$, $i = 1, \dots, n$, plot the values of $\frac{\mathbf{Z}_i}{n}$. See Fig. 4.

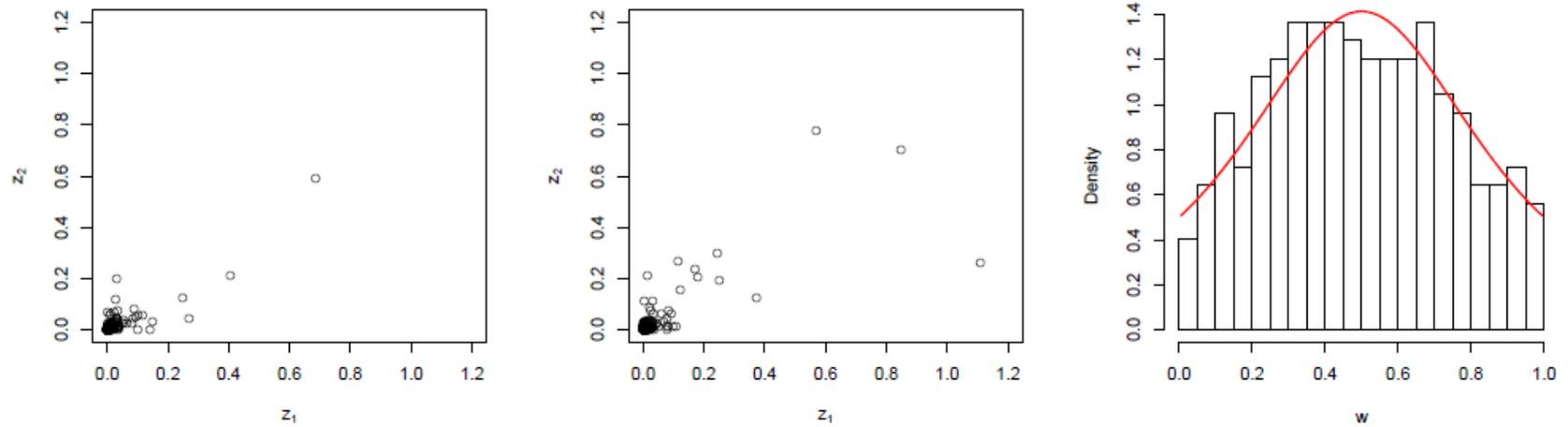


FIGURE 4: Scatterplots of z_i/n , where z_i are realizations from a bivariate logistic random vector and $n = 500$ (left) and $n = 5000$ (center). The right panel shows the angular density h along with a histogram of the angular components of the largest 5% of the realizations.

Regular Variation

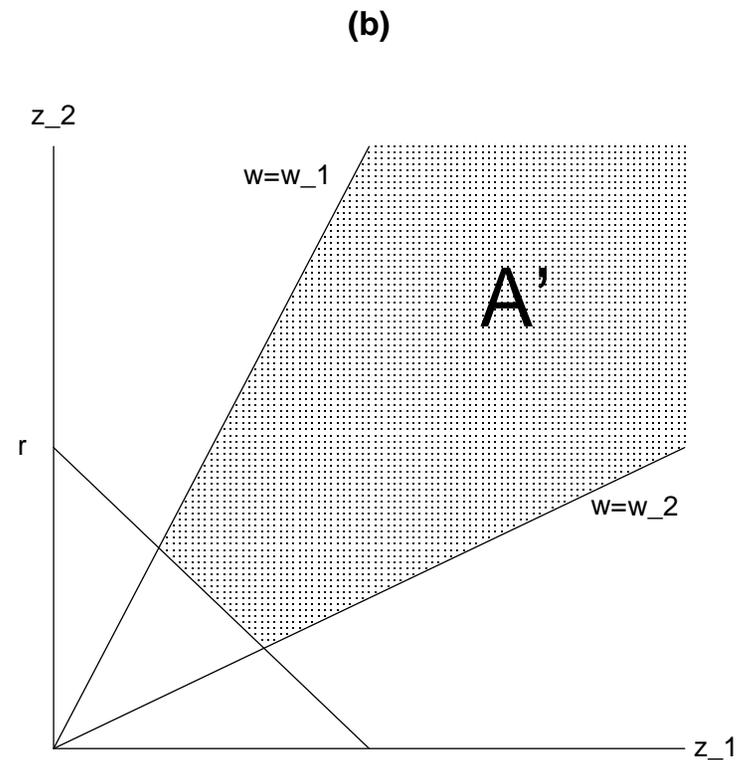
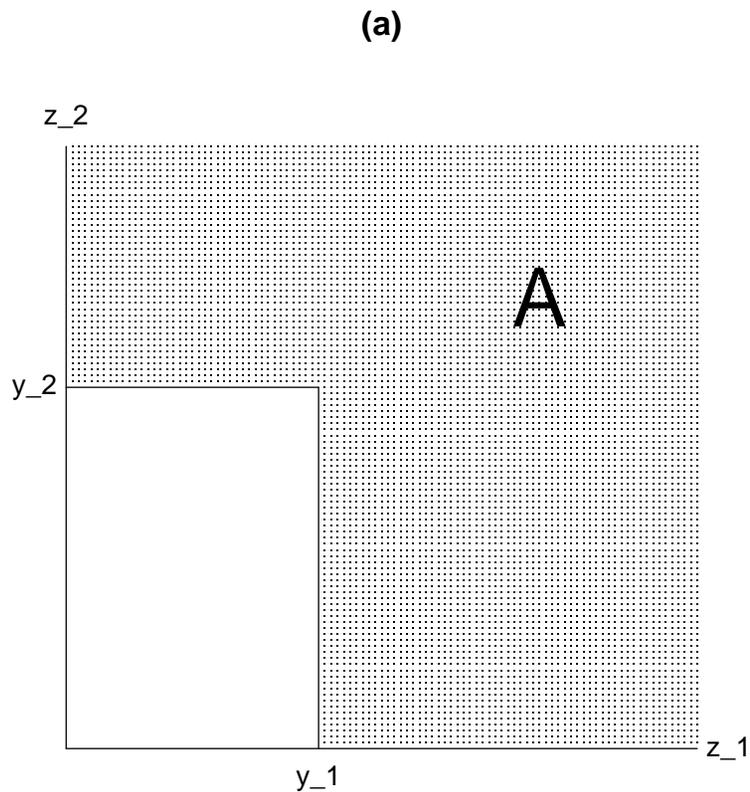
- Assume IID vectors $\mathbf{Z}_i = (Z_{i,1}, Z_{i,2})$ for $i = 1, \dots, n$. Generic name \mathbf{Z} . Let $M_{n,j} = \max(Z_{1,j}, \dots, Z_{n,j})$ for $j = 1, 2$.
- RV condition: for “continuity sets” A ,

$$nP \left(\frac{\mathbf{Z}}{\mathbf{a}_n} \in A \right) \xrightarrow{v} \nu(A)$$

where \xrightarrow{v} means vague convergence. In the unit Fréchet case, $\mathbf{a}_n = (n, n)$.

- Consider a set A of the form $\{(z_1, z_2) : z_1 > y_1 \text{ or } z_2 > y_2\}$ (see (a) on next figure)
- Then

$$\begin{aligned} P \left(\frac{M_{n,1}}{a_{n,1}} \leq y_1, \frac{M_{n,2}}{a_{n,2}} \leq y_1 \right) &= P^n \left(\frac{\mathbf{Z}}{\mathbf{a}_n} \notin A \right) \\ &\approx \left(1 - \frac{\nu(A)}{n} \right)^n \\ &\rightarrow e^{-\nu(A)}. \end{aligned}$$



Regular Variation Page 2

- If we consider the case when A is defined by (y_1, y_2) , we write instead $\nu(A) = V(\mathbf{y})$ which satisfies the scaling condition

$$\nu(cA) = c^{-1/\xi} \nu(A) \text{ (Unit Fréchet : } \xi = 1)$$

- Alternative: write in polar coordinates, $R = \|\mathbf{Z}\|$, $\mathbf{W} = \frac{\mathbf{Z}}{R}$.
- Could use any norm, but easiest is $\|\mathbf{Z}\| = |z_1| + |z_2|$.
- Consider set $A' = \{(R, \mathbf{W}), R > r, \mathbf{W} \in B\}$ for some $r > 0$ and angle set B (see (b) of figure). Then

$$\begin{aligned} nP \left(\frac{\mathbf{Z}}{\mathbf{a}_n} \in A' \right) &= nP \left(\frac{R}{a_n} > r, W \in B \right) \\ &\rightarrow r^{-1/\xi} H(B) \end{aligned}$$

where H is some measure on the space of \mathbf{W} .

Derivation of $V(\mathbf{y})$

- A is the set $(z_1 > y_1 \text{ or } z_2 > y_2)$
- Set $Z_1 = Rw$, $Z_2 = R(1 - w)$, the condition becomes $R > \min\left(\frac{y_1}{w}, \frac{y_2}{1-w}\right)$.
- So $V(\mathbf{y})$ is the same as

$$\begin{aligned}\nu(A) &= \int_0^\infty \int_0^\infty I(\mathbf{Z} \in A) d\nu(\mathbf{Z}) \\ &= \int_{w_1}^{w_2} \int_r^\infty I((Rw, R(1-w)) \in A) q R^{-2} dR dH(w) \\ &= \int_{w_1}^{w_2} \max\left(\frac{w}{y_1}, \frac{1-w}{y_2}\right) dH(w).\end{aligned}$$

- $H(w)$ is an arbitrary measure over $[0, 1]$ subject to

$$\int_0^1 w dH(w) = \int_0^1 (1-w) dH(w) = 1.$$

- This is true because $V(y, \infty) = V(\infty, y) = \frac{1}{y}$ for $0 < y < \infty$.

Examples of Santa Ana winds

- A weather condition defined by a combination of high winds and dry conditions — conducive to wildfires
- Example from 2003: Cedar Fire of 10/25/03. Another more extreme example was the Witch Fire of 2007.
- Data from 2003 shows combination of high windspeeds and low night time relative humidity.
- Define a *wind index* and a *dryness index* from these

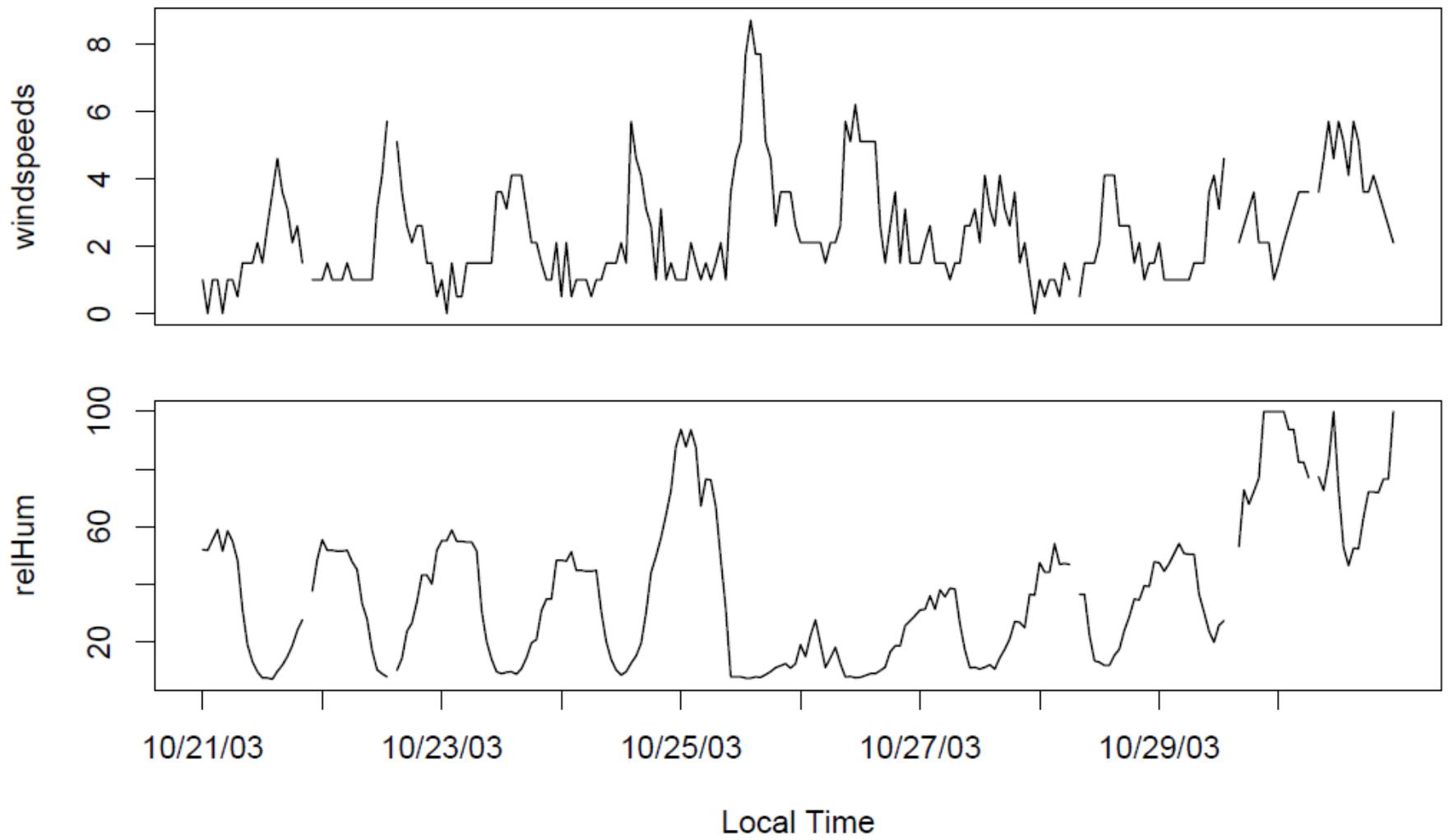


FIGURE 5: Time series of hourly windspeed (m/s) and relative humidity (%) for 10/21/2003 through 10/31/2003. The Cedar fire occurred on 10/25/2003.

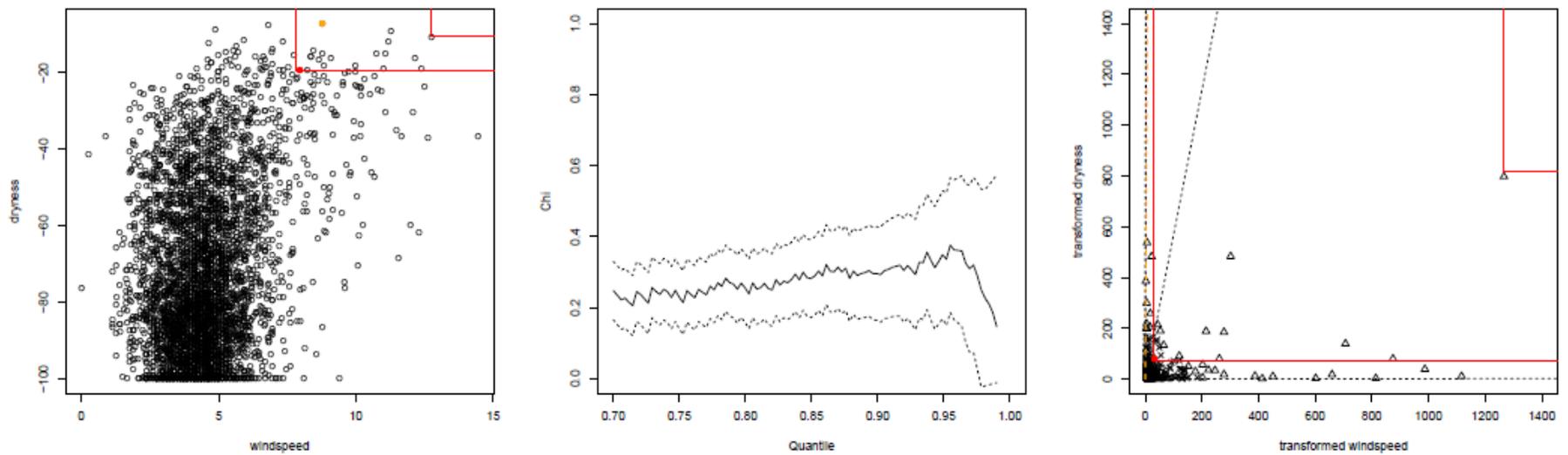


FIGURE 6: Left: scatterplot of daily summary values for windspeed and dryness. Center: plot of $\hat{\chi}$. Right: scatterplot of daily summaries after marginal transformation to unit Fréchet. The red solid lines region define risk regions \mathcal{R}_1 and \mathcal{R}_2 in the original scale and \mathcal{R}_1^* and \mathcal{R}_2^* in the Fréchet scale. The red dot corresponds to the day of the Cedar Fire, and the orange dot corresponds to the day of the Witch Fire. Dashed rays in the Fréchet scale plot point to events which lie outside the plot window.

Formulating Extreme Values Model

- Initial plot of untransformed data suggests positive dependence. Define two potential risk regions \mathcal{R}_1 and \mathcal{R}_2 respectively based on Cedar Fire and on largest combined value in the series (red boxes). Our *objective* is to estimate exceedance probabilities associated with these events
- Too extreme to estimate empirically, especially \mathcal{R}_2
- Dependence plot suggests $\chi \approx 0.3$ — “asymptotically dependent” case
- Transform to unit Fréchet margins — use GPD above 95th percentile and empirical CDF for the rest
- Risk regions now \mathcal{R}_1^* and \mathcal{R}_2^* . Illustrates extreme behavior but also hard to visualize (5 points outside picture, including Witch Fire)

Estimating Risk Region Probabilities

- Fit parametric model for H — suggests “mixture of betas” model (see histograms)
- First method: plug in parameter estimates. $\hat{p}_1 \approx 3.5 \times 10^{-3}$, $\hat{p}_2 \approx 1.7 \times 10^{-4}$.
- Second method: nonparametric, use scaling property. $\hat{p}_1 \approx 6.8 \times 10^{-3}$, $\hat{p}_2 \approx 1.5 \times 10^{-4}$.
- Confidence intervals by bootstrapping
- Compare with Gaussian copula — probability estimates this way are much higher
- No direct implication for climate change but could explore that using climate models

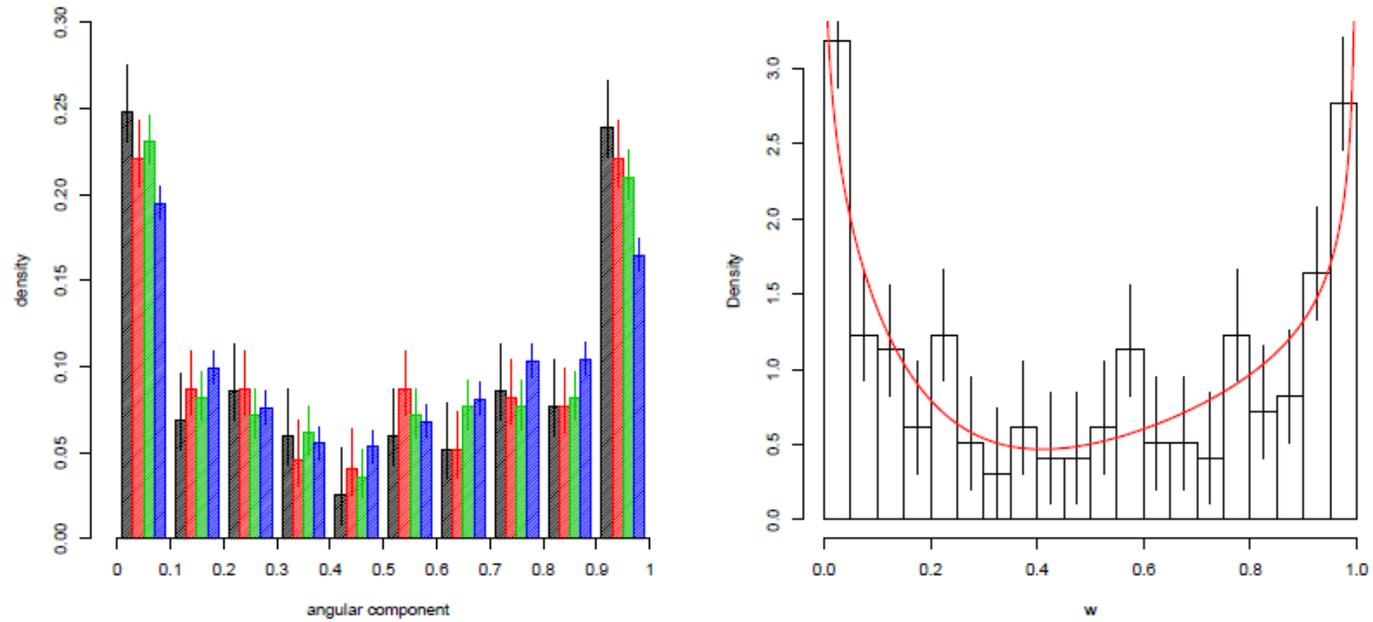


FIGURE 7: Left: histograms of the angular components of the events whose radial components exceed the .97 (black), .95 (red), .90 (green), and .75 (blue) quantiles. Right fitted parametric angular measure model of a mixture of two beta distributions.

Conclusions

- Univariate and bivariate cases
- Although parametric models have been used (GEV, GPD, parametric models for H) these are fitted only to the tails of the distribution (“letting the tail speak for itself”)
- Parameter estimation is not straightforward but we have used both MLE and Bayesian methods. Estimation of uncertainty is critical.
- In climate examples, more complex analyses possible in principle by combining observational and climate model output. This has applications to “extreme event attribution” and projection of future extreme event probabilities.