Statistical Methods for Analyzing Climate Extremes

Whitney Huang
CLEMSON UNIVERSITY

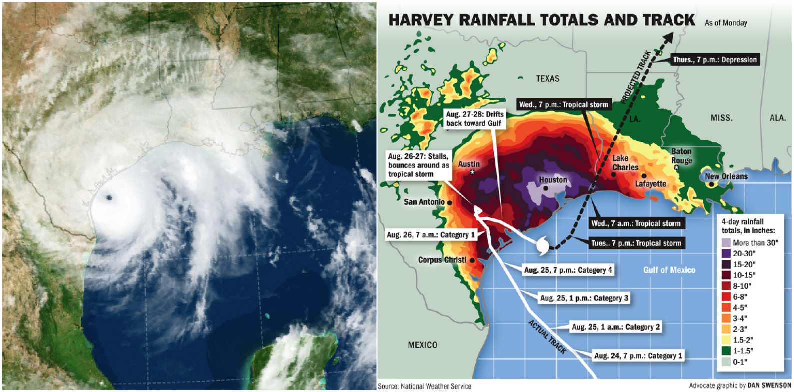
Conference on Mathematics of Planet Earth
Pittsburgh, PA, July 14, 2022

This tutorial is (largely) based on some overview talks given by Dan Cooly (CSU) and Anthony Davison (EPFL)

Agenda

Outline

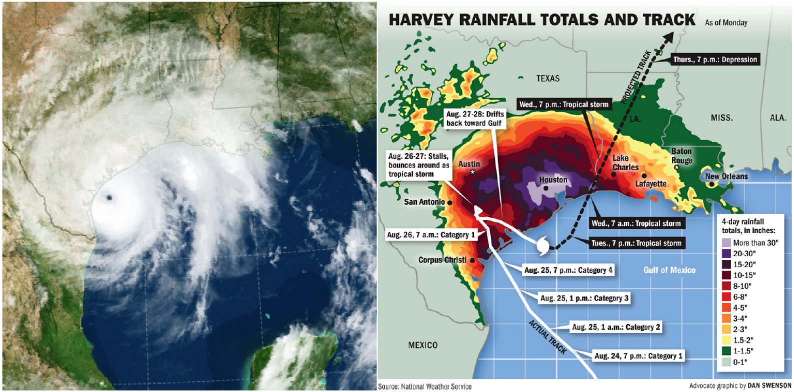
Extreme Rainfall During Hurricane Harvey



Source: NASA (Left); National Weather Service (Right)

- The highest total rainfall was 60.58 inches near Nederland, TX
- **Annual average** rainfall for Pittsburgh: 38.3 inches

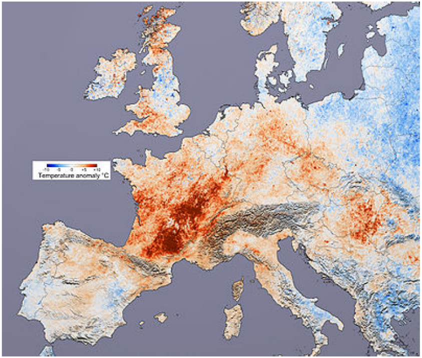
Extreme Rainfall During Hurricane Harvey



Source: NASA (Left); National Weather Service (Right)

- The highest total rainfall was 60.58 inches near Nederland, TX
- **Annual average** rainfall for Pittsburgh: 38.3 inches

Some Examples of Weather/Climate Extremes

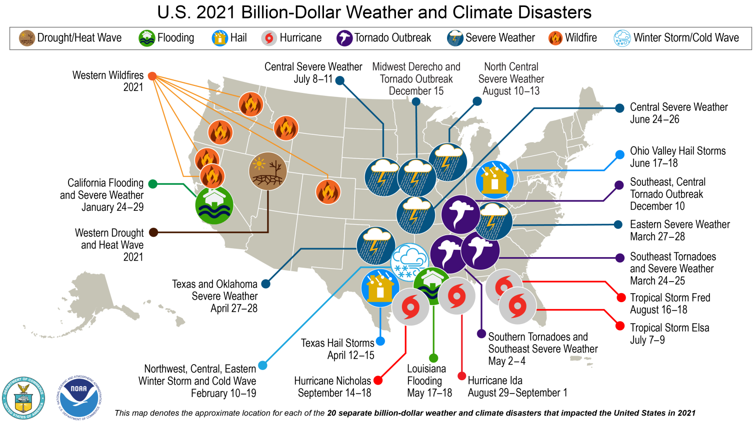


- **Heat wave**: The 2003 European heat wave led to the hottest summer on record in Europe since 1540 that resulted in at least 30,000 deaths

- **Storm Surge**: Hurricane Katrina produced the highest storm surge ever recorded (27.8 feet) on the U.S. coast

Why Study Climate Extremes?

Although infrequent, extremes usually have large impact.



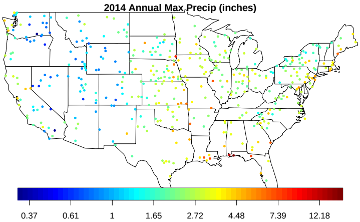
Source: National Oceanic and Atmospheric Administration

Goal: to quantify the tail behavior $\Rightarrow$ often requires extrapolation.

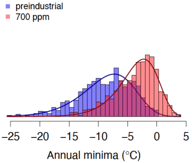
Some Scientific Questions

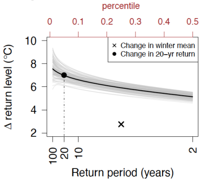
- How to estimate the magnitude of extreme events (e.g. 100-year rainfall)?

- How extremes vary in space?



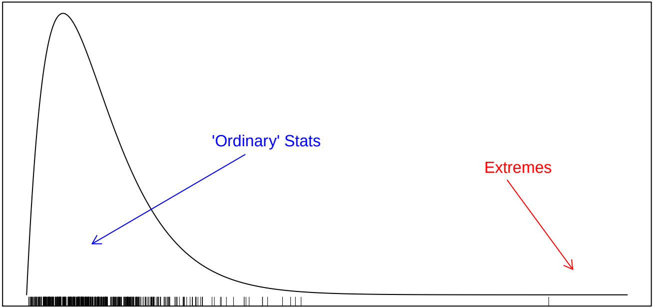
- How extremes change in future climate conditions?





Outline

Usual vs Extremes



	Target	Theory	Distribution
Ordinary Stats	bulk distribution	CLT	Normal
Extreme Stats	tail distribution(s)	?	?

Probability Framework for the Sample Maximum

Let $X_1, \cdots, X_n \overset{i.i.d.}{\sim} F$ and define $M_n = \max\{X_1, \cdots, X_n\}$
Then the distribution function of M_n is

$$\begin{aligned}\mathbb{P}(M_n \leq x) &= \mathbb{P}(X_1 \leq x, \cdots, X_n \leq x) \\ &= \mathbb{P}(X_1 \leq x) \times \cdots \times \mathbb{P}(X_n \leq x) = F^n(x)\end{aligned}$$

Remark

$$F^n(x) \xlongequal{n \to \infty} \begin{cases} 0 & \text{if } F(x) < 1 \\ 1 & \text{if } F(x) = 1 \end{cases}$$

$\Rightarrow$ the limiting distribution is degenerate.

Asymptotic: Classical Limit Laws

Recall the Central Limit Theorem (CLT):

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} N(0,1),$$

where $S_n = \sum_{i=1}^{n} X_i$

$\Rightarrow$ rescaling is the key to obtain a non-degenerate distribution

Question: Can we get the limiting distribution of

$$\frac{M_n - b_n}{a_n}$$

for suitable sequence $\{a_n\} > 0$ and $\{b_n\}$?

Asymptotic: Classical Limit Laws

Recall the Central Limit Theorem (CLT):

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} N(0,1),$$

where $S_n = \sum_{i=1}^{n} X_i$

$\Rightarrow$ rescaling is the key to obtain a non-degenerate distribution

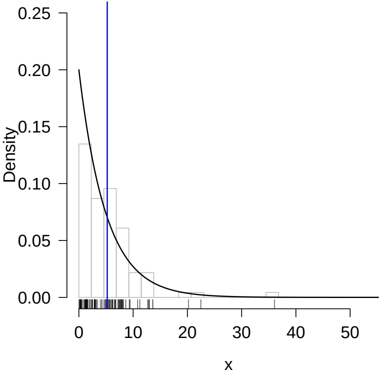
Question: Can we get the limiting distribution of

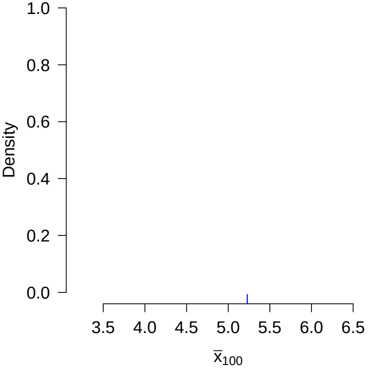
$$\frac{M_n - b_n}{a_n}$$

for suitable sequence $\{a_n\} > 0$ and $\{b_n\}$?

CLT in Action

1. Generate 100 (n) random numbers from an Exponential distribution (population distribution)
2. Compute the sample mean of these 100 random numbers
3. Repeat this process 120 times





Extremal Types Theorem [Fisher–Tippett 1928, Gnedenko 1943]

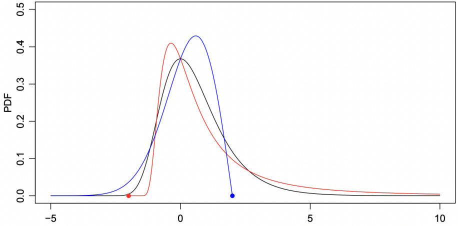
Define $M_n = \max\{X_1, \cdots, X_n\}$ where $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} F$. If $\exists\, a_n > 0$ and $b_n \in \mathbb{R}$ such that, as $n \to \infty$, if

$$\mathbb{P}\left((M_n - b_n)/a_n \leq x\right) \xrightarrow{d} \mathsf{G}(x)$$

then G must be the same type of the following form:

$$\mathsf{G}(x; \mu, \sigma, \xi) = \exp\left\{ -\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{\frac{-1}{\xi}} \right\}$$

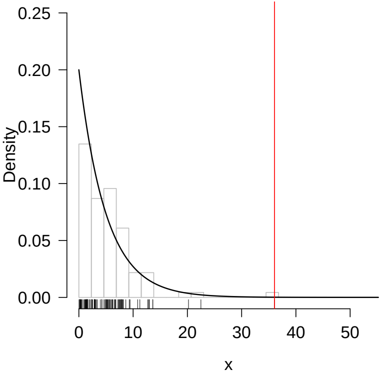
where $x_+ = \max(x, 0)$ and $G(x)$ is the distribution function of the generalized extreme value distribution (GEV(μ, σ, ξ)), where μ and σ are location and scale parameters, and ξ is the shape parameter

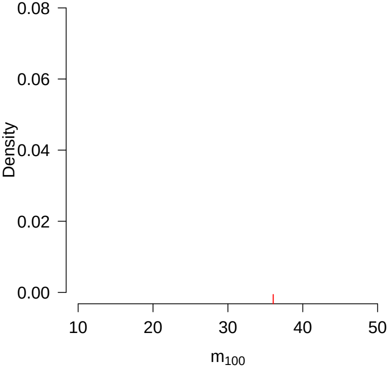


- $\xi > 0$: Fréchet (heavy-tail)
- $\xi = 0$: Gumbel (light-tail)
- $\xi < 0$: reversed Weibull (short-tail)

Extremal Types Theorem in Action

1. Generate 100 (n) random numbers from an Exponential distribution (population distribution)
2. Compute the sample maximum of these 100 random numbers
3. Repeat this process 120 times





Max-Stability and GEV

Definition

A distribution G is said to be max-stable if

$$G^k(a_k x + b_k) = G(x), \qquad k \in \mathbb{N}$$

for some constants $a_k > 0$ and b_k

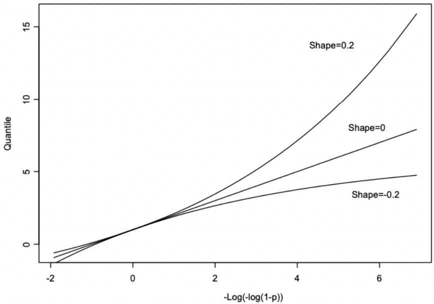
- Taking powers of a distribution function results only in a change of location and scale
- A distribution is max-stable $\iff$ it is a GEV distribution

Quantiles and Return Levels

- Quantiles of GEV

$$G(m_p) = \exp\left\{ -\left[1+\xi\left(\frac{m_p-\mu}{\sigma}\right)\right]_+^{\frac{-1}{\xi}} \right\} = 1-p$$

$$\Rightarrow m_p = \mu - \frac{\sigma}{\xi}\left[1 - \{-\log(1-p)^{-\xi}\}\right] \qquad 0<p<1$$



- In the extreme value terminology, m_p is the return level associated with the return period $\frac{1}{p}$

Outline

Statistical Practice

Assume n is large enough so that

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) \approx \exp\left(-[1+\xi x]^{-1/\xi}\right)$$

$$\Rightarrow \mathbb{P}(M_n \leq y) \approx \exp\left(-\left[1+\xi\left(\frac{y-b_n}{a_n}\right)^{-1/\xi}\right]\right)$$

$$:= \exp\left(-\left[1+\xi\left(\frac{y-\mu}{\sigma}\right)\right]^{-1/\xi}\right)$$

Then, we have a three-parameter estimation problem. μ, σ, ξ can be estimated via maximum likelihood[1]

[1] Probability weighted moments/L-moments and Bayesian methods can also be used to carry out parameter estimation

Maximum Likelihood Estimation

Let $M_1, \cdots, M_k \overset{\text{iid}}{\sim} \text{GEV}$, then log-likelihood for (μ, σ, ξ) when $\xi \neq 0$ is

$$\ell(\mu, \sigma, \xi) = -m \log \sigma - (1 + 1/\xi) \sum_{i=1}^{k} \log \left[1 + \xi \left(\frac{m_i - \mu}{\sigma} \right) \right] - \sum_{i=1}^{k} \left[1 + \xi \left(\frac{m_i - \mu}{\sigma} \right) \right]^{-1/\xi},$$

provided that $1 + \xi(\frac{m_i - \mu}{\sigma}) > 0$, for $i = 1, \cdots, k$.

When $\xi = 0 \rightarrow$ use the Gumbel limit of the GEV

Maximum likelihood estimate (MLE) is obtained by (numerically) maximization of log-likelihood shown above

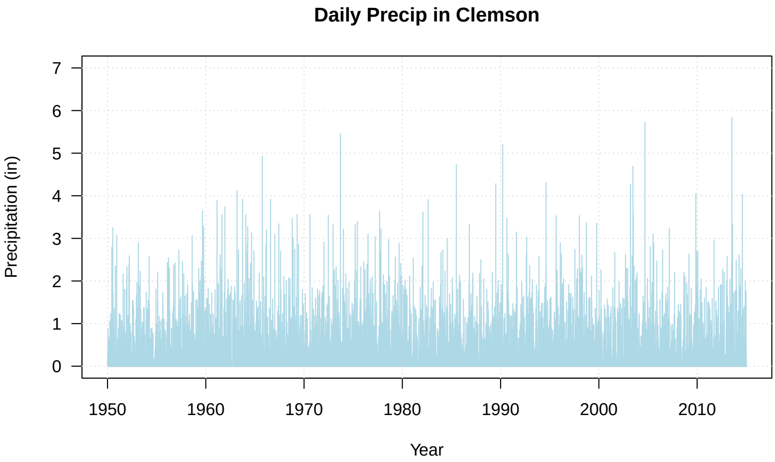
Uncertainty Quantification for GEV Estimation

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

Two method:

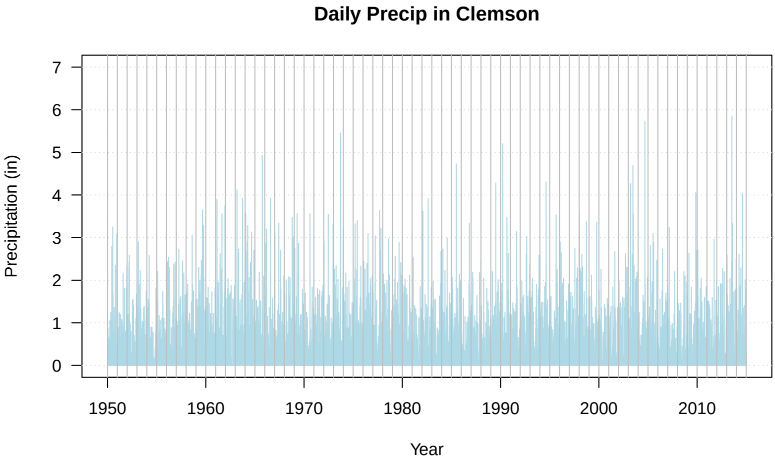
- Delta method
  - $+$: easy to compute with a closed form expression
  - -: symmetric confidence interval is not realistic (especially for long return levels)
- Profile likelihood method
  - $+$: can allow for asymmetric confidence intervals
  - -: need to compute numerically

Clemson Daily Precipitation [Data Source: USHCN]



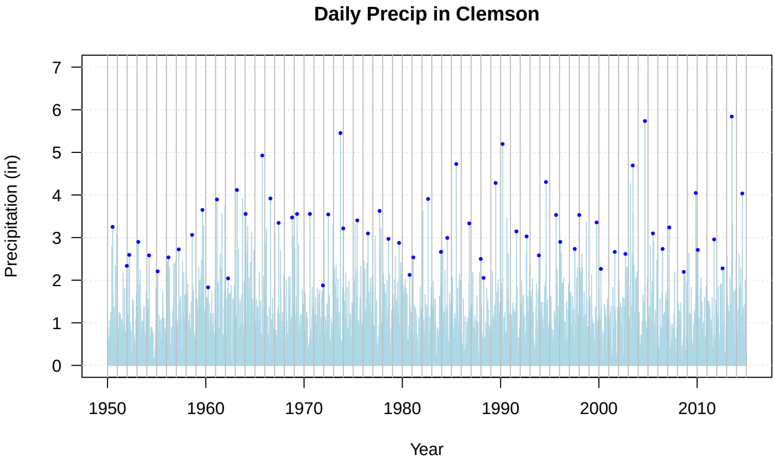
Block-Maximum Method [Gumbel 1958]

1. Determine the block size and extract the block maxima



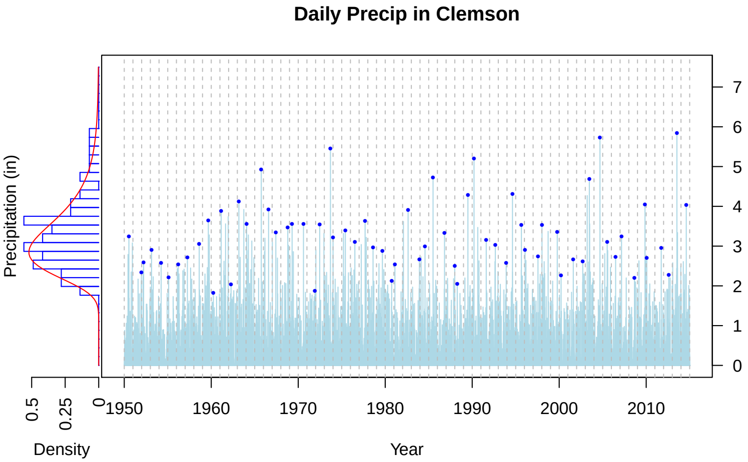
Block-Maximum Method [Gumbel 1958]

1. Determine the block size and extract the block maxima



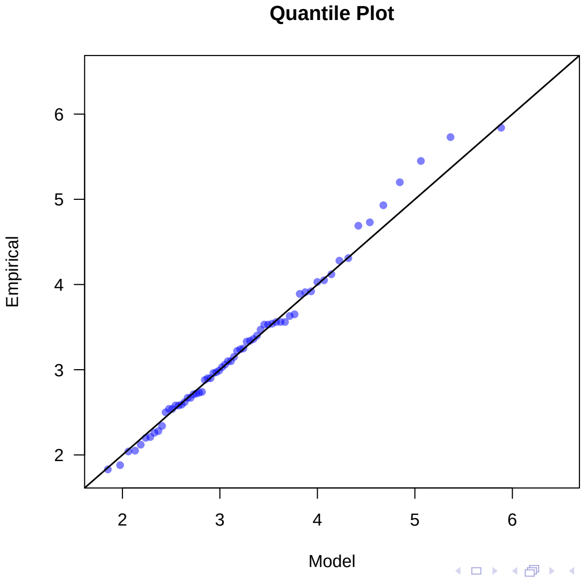
Block-Maximum Method [Gumbel 1958]

2. Fit the GEV to the maximal and assess the fit



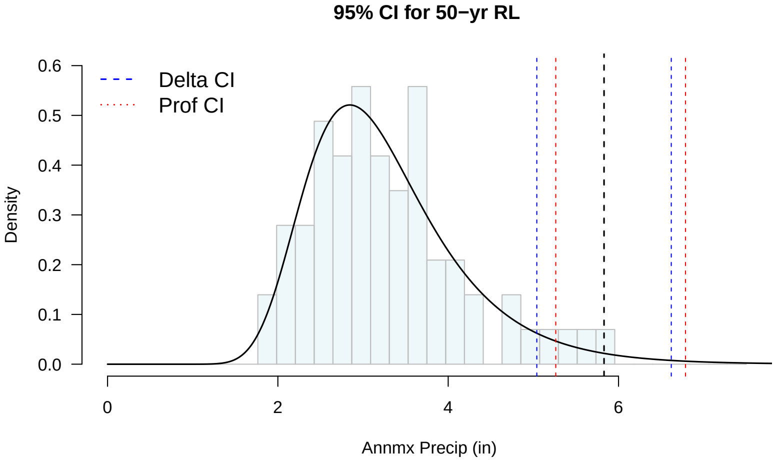
Block-Maximum Method [Gumbel 1958]

2. Fit the GEV to the maximal and assess the fit

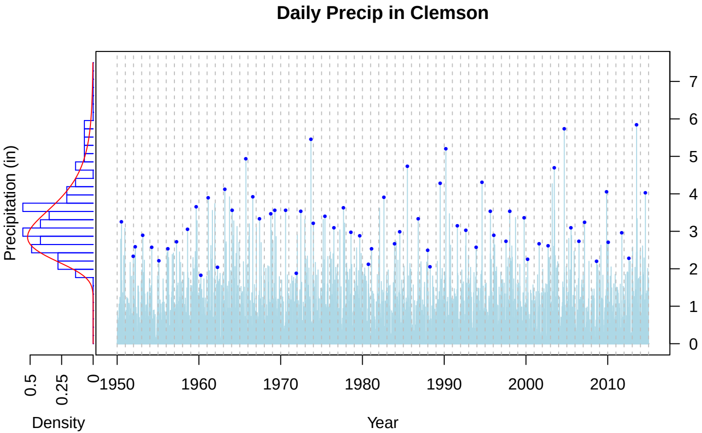


Block-Maximum Method [Gumbel 1958]

3. Perform inference for return levels, probabilities, etc.



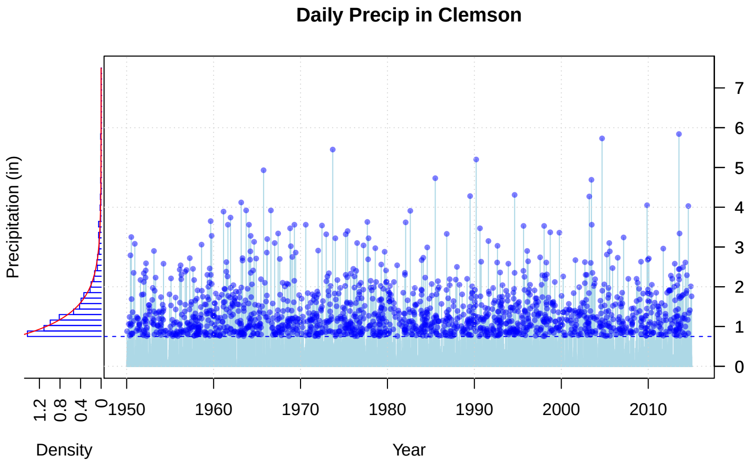
Can We Get More Data for Extreme Value Analysis?



Question: Can we use data more efficiently to infer extremes?

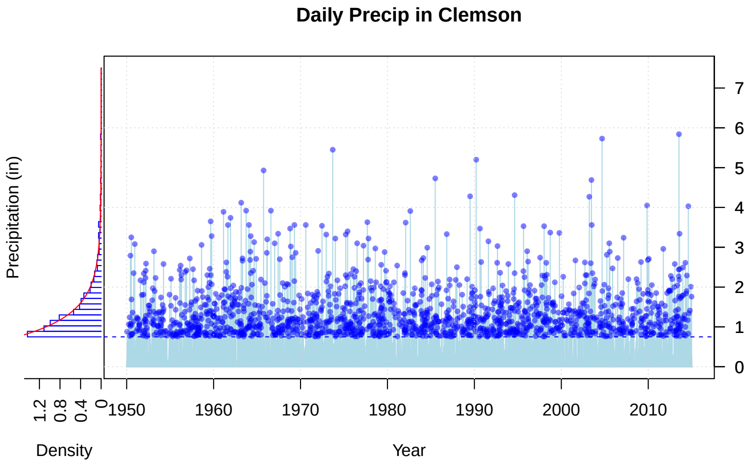
Threshold-Exceedance Method [Davison & Smith 1990]

1. Select a "sufficiently large" threshold u, extract the exceedances



Threshold-Exceedance Method [Davison & Smith 1990]

2. Fit an appropriate model to exceedances



Generalized Pareto Distribution for Exceedances

If $M_n = \max_{i=1,\cdots,n} X_i$ (for a large n) can be apprximated by a GEV(μ, σ, ξ), then for sufficently large u,

$$\begin{aligned}
\mathbb{P}(X_i > x + u | X_i > u) &= \frac{n\mathbb{P}(X_i > x + u)}{n\mathbb{P}(X_i > u)} \\
&\to \left(\frac{1 + \xi \frac{x+u-b_n}{a_n}}{1 + \xi \frac{u-b_n}{a_n}} \right)^{\frac{-1}{\xi}} \\
&= \left(1 + \frac{\xi x}{a_n + \xi(u - b_n)} \right)^{\frac{-1}{\xi}}
\end{aligned}$$

$\Rightarrow$ Survival function of generalized Pareto distribution (GPD)

Pickands–Balkema–de Haan Theorem (1974, 1975)

If $M_n = \max_{1 \leq i \leq n}\{X_i\} \approx \text{GEV}(\mu, \sigma, \xi)$, then, for a "large" u (i.e., $u \to x_F = \sup\{x : F(x) < 1\}$),

$$\mathbb{P}(X > u) \approx \frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]^{\frac{-1}{\xi}}$$

$F_u = \mathbb{P}(X - u < y | X > u)$ is well approximated by the generalized Pareto distribution (GPD). That is:

$$F_u(y) \xrightarrow{d} H_{\tilde{\sigma},\xi}(y) \qquad u \to x_F$$

where

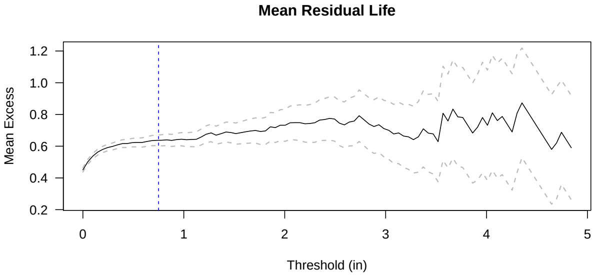
$$H_{\tilde{\sigma},\xi}(y) = \begin{cases} 1 - (1 + \xi y/\tilde{\sigma})^{-1/\xi} & \xi \neq 0; \\ 1 - \exp(-y/\tilde{\sigma}) & \xi = 0. \end{cases}$$

and $\tilde{\sigma} = \sigma + \xi(u - \mu)$

How to Choose the Threshold?

Bias–variance tradeoff:

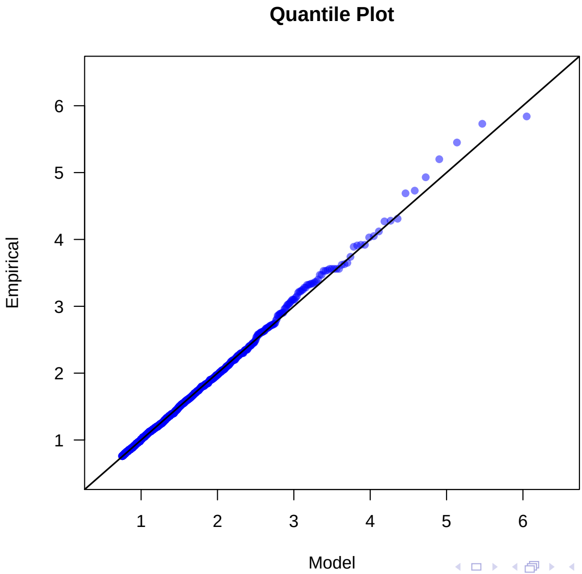
- Threshold too low $\Rightarrow$ bias because of the model asymptotics being invalid
- Threshold too high $\Rightarrow$ variance is large due to few data points



Task: To choose a u_0 s.t. the Mean Residual Life curve behaves linearly $\forall u > u_0$

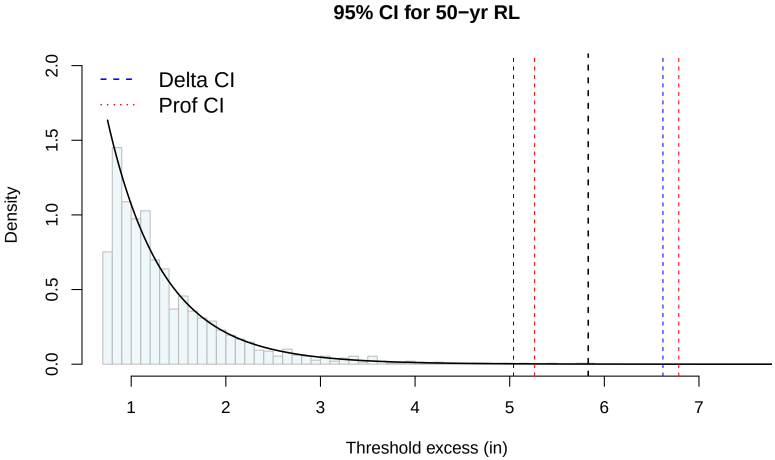
Threshold-Exceedance Method [Davison & Smith 1990]

2. Fit an appropriate model to exceedances and assess the fit



Threshold-Exceedance Method [Davison & Smith 1990]

3. Perform inference for return levels, probabilities, etc



Outline

Temporal Dependence

Question: Is the GEV still the limiting distribution for block maxima of a stationary (but not independent) sequence $\{X_i\}$?

Answer: Yes, as long as mixing conditions hold. (Leadbetter et al., 1983)

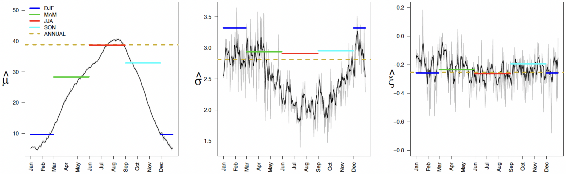
What does this mean for inference?

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

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

Modeling Non-Stationary Extremes: Seasonality and Long-Term Trend

- $M_t \sim \text{GEV}(\mu(t), \sigma(t), \xi(t))$



- Typically assume fairly simple structure for $\mu(t)$ and $\sigma(t)$,

$$\text{e.g. } \mu(t) = \mu_0 + \mu_1 t,$$

  and $\xi(t)$ be a constant

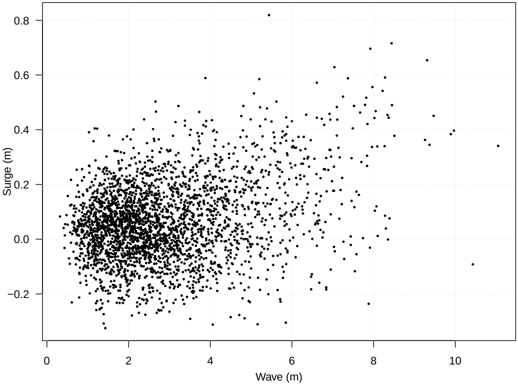
- $\mu(t)$ and $\sigma(t)$ could depend on some **physically-informed factors** (e.g. Clausius-Clapeyron precipitation-temperature scaling)

Remarks on Univariate Extremes

- To estimate the tail, EVT uses only extreme observations
- Shape parameter ξ is extremely important but hard to estimate
- Threshold exceedance approaches allow the user to retain more data than block-maximum approaches, thereby reducing the uncertainty with parameter estimates
- Temporal dependence in the data is more of an issue in threshold exceedance models. One can either decluster, or alternatively, adjust inference

Outline

Wave–Surge Data at Newlyn, UK

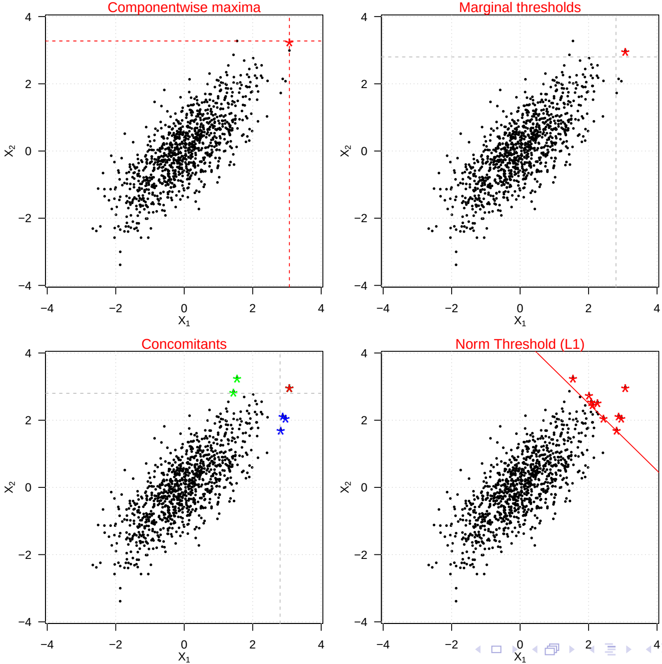


- Scientific motivation: many extremal problems are intrinsically multivariate

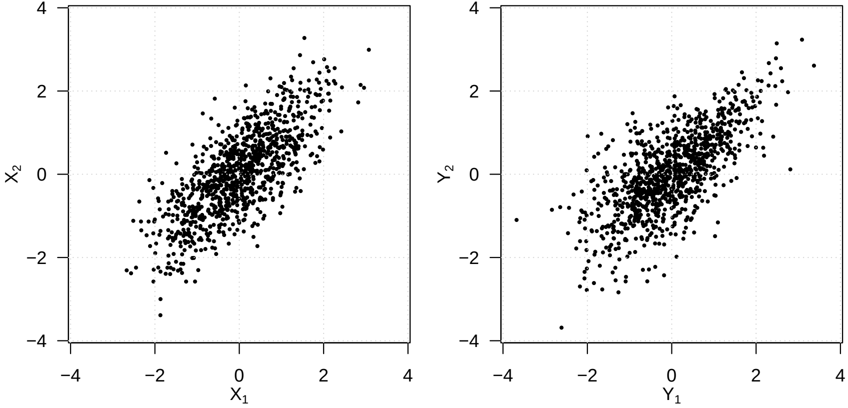
- Statistical motivation: uncertainty may be reduced by combining information from several sources

Challenge: what is "extreme" in two or more dimensions?

What is a Multivariate Extreme?

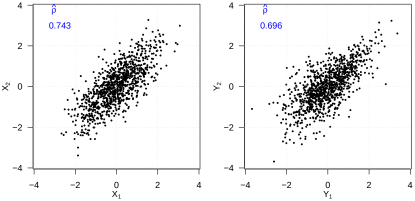


Describing Tail Dependence



A central aim of multivariate extremes is to describe tail dependence

Describing Tail Dependence (Cont'd)

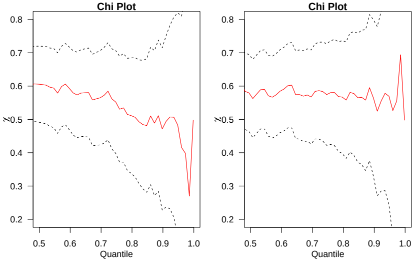


Correlation Coefficient

$$\rho = \frac{\mathbb{E}\left(\left(X_1 - \mu_{X_1}\right)\left(X_2 - \mu_{X_2}\right)\right)}{\sqrt{\mathrm{Var}(X_1)\mathrm{Var}(X_2)}}$$

ρ measures "spread from center"

$\Rightarrow$ does not focus on extremes



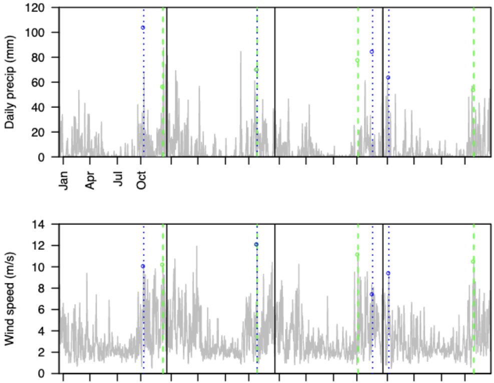
Upper tail dependence parameter:

$$\chi = \lim_{u \to 1} \chi(u)$$
$$= \mathbb{P}\left(F_{X_1}(X_1) > u | F_{X_2}(X_2) > u\right)$$

Outline

Componentwise Maxima

Let $\{\boldsymbol{X}_i = (X_{1,i}, \cdots, X_{d,i})\} \in \mathbb{R}^d$ are i.i.d. random vectors. Let's first consider **componentwise maxima**:
$\boldsymbol{M}_n = (M_{1,n}, \cdots, M_{d,n})^T$, where $M_{j,n} = \max_{i=1}^n X_{j,i}$



- $\boldsymbol{M}_n$ is not necessarily one the the original observations
- Univariate GEV still apply in each margin
- Need to model the interdependence across $\{M_{j,n}\}_{j=1}^d$

Multivariate Extreme Value Theorem

Let $\boldsymbol{X}_1, \cdots, \boldsymbol{X_n} \overset{\text{iid}}{\sim} \boldsymbol{F}$ where $\boldsymbol{X}_i \in \mathbb{R}^d$, if

$$\frac{\max_{i \le i \le n} \mathbf{X}_i - \mathbf{b}_n}{\mathbf{a}_n} = \left(\frac{\bigvee_{i=1}^n X_{1,i} - b_{1,n}}{a_{1,n}}, \cdots, \frac{\bigvee_{i=1}^n X_{d,i} - b_{d,n}}{a_{d,n}} \right) \xrightarrow{d} \mathbf{G}$$

then $\mathbf{G}$ must be a multivariate extreme value distribution

- Each marginal (approximately) follows a GEV distribution (i.e., $M_j \approx \mathrm{GEV}(\mu_j, \sigma_j, \xi_j)$)
- No limiting parametric family exists for describing extremal dependence, i.e., the interdependence across $\{M_{j,n}\}_{j=1}^d$
- Marginals and dependence are typically handled separately $\Rightarrow$ "Copula-like" approach

Multivariate GEV

- Transform each marginal to a common marginal distribution, e.g., unit Fréchet $(\mathrm{GEV}(1,1,1))$

$$\tilde{M}_j = \left[1 + \xi_j \left(\frac{M_j - \mu_j}{\sigma_j}\right)\right]^{\frac{1}{\xi_j}}$$

with $\tilde{G}(m) = \exp(m^{-1})$

- Then

$$\mathbb{P}(\tilde{M}_1 \leq \tilde{m}_1, \cdots, \tilde{M}_d \leq \tilde{m}_d) = \exp\left(-V\left(\tilde{m}_1, \cdots, \tilde{m}_d\right)\right),$$

where $V : \mathbb{R}_+^d \mapsto \mathbb{R}_+$ is the exponent measure that characterizes the extremal dependence

- In practice, modeling usually involves fitting a parametric family of V

The Logistic Model

$$G(m_1, m_2) = \exp\left\{-\left(m_1^{-\frac{1}{\alpha}} + m_2^{-\frac{1}{\alpha}}\right)^{\alpha}\right\},$$

where m_1 and $m_2 > 0$, $\alpha \in (0, 1)$

- $\alpha \to 1$ corresponds to independent
- $\alpha \to 0$ corresponds to perfectly dependent
- This model is symmetric $\Rightarrow$ the variables are exchangeable, that is, M_1 depends on M_2 to exactly the same degree that M_2 depends on M_1

The Bilogistic Model

$$G(m_1, m_2) = \exp\left\{m_1\gamma^{1-\alpha} + m_2(1-\gamma)^{1-\beta}\right\},$$

where $0 < \alpha < 1$, $0 < \beta < 1$, and γ is the solution of:

$$(1-\alpha)m_1(1-\gamma)^{\beta} = (1-\beta)m_2\gamma^{\alpha}$$

- Independence is obtained when $\alpha = \beta \to 1$ and when one of α or β is fixed and the other approaches 1
- When $\alpha = \beta$ the model reduces to the logistic model
- The value of $\alpha - \beta$ determines the extent of asymmetry in the dependence structure

Parameter Estimation

Models (e.g., Logistic or bilogistic) can be fitted by maximum likelihood estimation, either:

- **two steps**: marginal estimates GEV marginals followed by dependence function, or
- **one step**: estimates marginals and dependence structure simultaneously

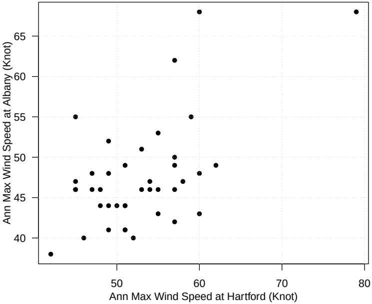
The probability density function is

$$g(m_1, m_2) = \{V_{m_1} V_{m_2} - V_{m_1 m_2}\} \exp(-V(m_1, m_2)),$$

where V_{m_1}, V_{m_2}, and $V_{m_1 m_2}$ denote partial derivatives of V.

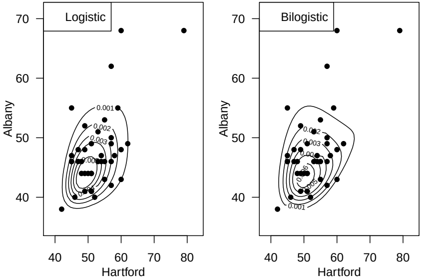
With the (transformed) data $(\tilde{m}_{1,i}, \tilde{m}_{2,i})_{i=1}^{m}$, one can maximze the log-likelihood $\ell(\theta) = \sum_{i=1}^{m} \log g(\tilde{m}_{1,i}, \tilde{m}_{2,i})$ to obtain maximum likelihood estimates and standard errors

Annual Maximum Wind Speeds at Albany and Hartford



Marginal Parameter Estimates

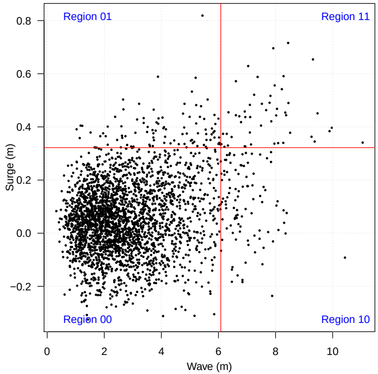
Model	Logistic	Bilogisitc
$\hat{\mu}_{\text{Hartford}}$	49.97 (0.87)	50.08 (0.84)
$\hat{\sigma}_{\text{Hartford}}$	5.04 (0.64)	5.07 (0.62)
$\hat{\xi}_{\text{Hartford}}$	0.01 (0.09)	-0.04 (0.05)
$\hat{\mu}_{\text{Albany}}$	44.58 (0.77)	44.64 (0.87)
$\hat{\sigma}_{\text{Albany}}$	4.34 (0.57)	4.40 (0.42)
$\hat{\xi}_{\text{Albany}}$	0.08 (0.11)	0.08 (0.12)



Dependence Parameter Estimates

Model	Logistic	Bilogistic
$\hat{\alpha}$	0.71 (0.10)	0.10 (0.00)
$\hat{\beta}$	NA	0.90 (0.05)
log-lik.	-246.07	-244.89

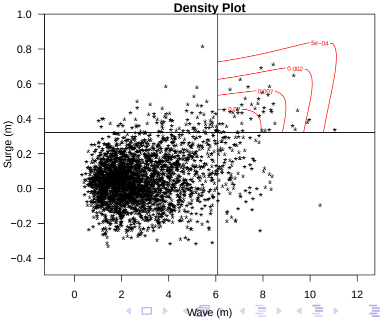
Bivariate Threshold Excess Model



$$g(x_1, x_2) =$$

$$\begin{cases} \dfrac{\partial^2 G}{\partial x_1 \partial x_2}\Big|_{(x_1,x_2)} & \text{if } (x_1, x_2) \in \text{Region 11;} \\ \dfrac{\partial G}{\partial x_1}\Big|_{(x_1,u_{x_2})} & \text{if } (x_1, x_2) \in \text{Region 10;} \\ \dfrac{\partial G}{\partial x_2}\Big|_{(u_{x_1},x_2)} & \text{if } (x_1, x_2) \in \text{Region 01;} \\ G(u_{x_1}, u_{x_2}) & \text{if } (x_1, x_2) \in \text{Region 00.} \end{cases}$$

Model	Logistic	Bilogistic
$\hat{\sigma}_{u_{\text{Wave}}}$	1.26 (0.13)	1.27 (0.13)
$\hat{\xi}_{\text{Wave}}$	-0.13 (0.07)	-0.14 (0.07)
$\hat{\sigma}_{u_{\text{Surge}}}$	0.09 (0.01)	0.09 (0.01)
$\hat{\xi}_{\text{Surge}}$	0.01 (0.09)	0.01 (0.08)
$\hat{\alpha}$	0.76 (0.03)	0.79 (0.05)
$\hat{\beta}$	NA	0.73 (0.07)
log-lik.	-1018.04	-1017.90



Asymptotic Dependence and Independence

One key problem with using limit distributions for multivariate extremes is that they force one of two possibilities:

- Asymptotic Dependence (AD): extremes occur with a dependence structure which conforms to an extreme value distribution
  $\rightarrow \chi = \lim_{u \to 1} \mathbb{P}\left(F_{X_1}(X_1) > u | F_{X_2}(X_2) > u\right) > 0$
- Asymptotic Independence (AI): extremes occur independently in the different margins
  $\rightarrow \chi = \lim_{u \to 1} \mathbb{P}\left(F_{X_1}(X_1) > u | F_{X_2}(X_2) > u\right) = 0$

Multivariate extreme value models interpret AI as exactly independent. **However**, AI is often suggested by the data, and yet quite strong dependence is still present at high levels (e.g., Gaussian copula with $\rho \neq 1$)

Outline

Conditional Extreme Value (CEV) Models [Heffernan & Tawn, 04]

Models the conditional distribution by assuming a **parametric** location-scale form after marginal transformation

- **Marginal modeling**:
  1. Estimate marginal distributions of Y and X
  2. Transform $(Y, X)^T$ to Laplace marginals $(\tilde{Y}, \tilde{X})^T$
- **Dependence modeling**:

  Assume for large u,

$$\left[\frac{\tilde{Y} - a\left(\tilde{X}\right)}{b\left(\tilde{X}\right)} \leq z | \tilde{X} > u\right] \sim G(z),$$

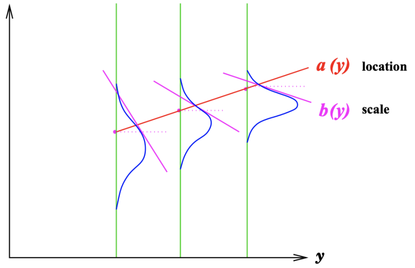
where $a(x) = \alpha x$ and $b(x) = x^{\beta}$, $\alpha \in [-1, 1]$, $\beta \in (-\infty, 1)$

A cartoon illustration of the CEV dependence modeling

Assume for large u,

$$\left[\frac{\tilde{Y} - a\left(\tilde{X}\right)}{b\left(\tilde{X}\right)} \leq z | \tilde{X} > u\right] \sim G(z),$$

where $a(x) = \alpha x$ and $b(x) = x^{\beta}$, $\alpha \in [-1, 1]$, $\beta \in (-\infty, 1)$



Source: Heffernan's slides given at the Interface 2008 Symposium

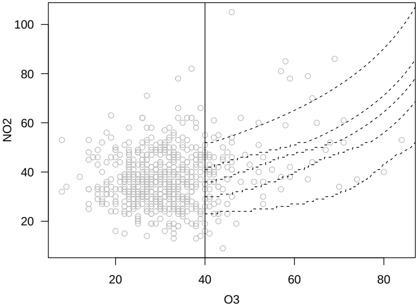
- $\tilde{Y} = \alpha \tilde{X} + \tilde{X}^{\beta} Z$,
  $\Rightarrow Z = \frac{\tilde{Y} - \alpha \tilde{X}}{\tilde{X}^{\beta}} \sim G$
- α and β are estimated by making a parametric assumption of $\tilde{Y}$
- G estimated nonparametrically

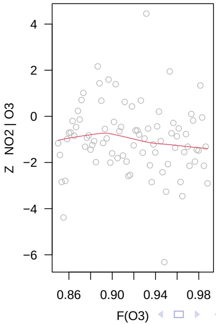
Modeling "Summer" O_3 and NO_2 [Heffernan & Tawn, 2004]

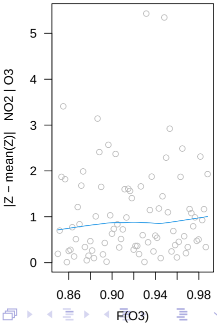
- Apr.-July, 1994-1998 daily maximum air pollution data from Leeds, U.K.

- Need to choice several thresholds: one for each marginal and one for the dependence structure

- $\tilde{Y} = \alpha \tilde{X} + \tilde{X}^{\beta} Z,$
  $\Rightarrow Z = \frac{\tilde{Y} - \alpha \tilde{X}}{\tilde{X}^{\beta}} \sim G$
  $\hat{\alpha} = 0.77, \ \hat{\beta} = 0.188$





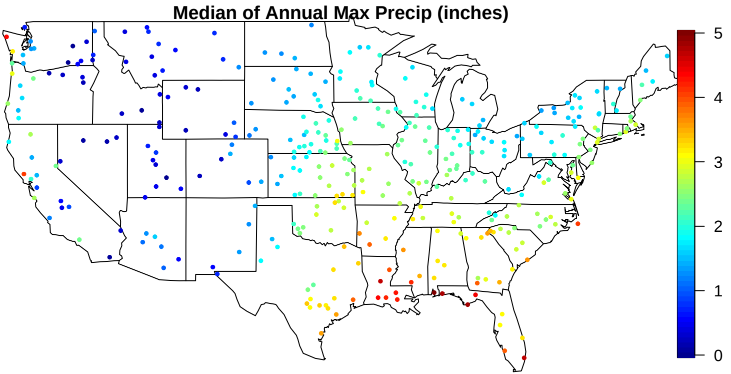


Remarks on Multivariate Extreme Value Analysis

- Definition of a multivariate extreme is not obvious
- Tail dependence NOT summarized with correlations, we looked at the upper tail dependence parameter χ
- Methodologies exist for both block maxima and threshold exceedance approaches
- Conditional extreme value models for asymptotic independence cases

Outline

Climate and Weather Spatial Effects



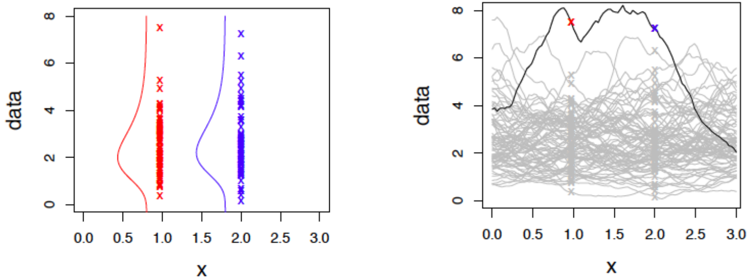
"*Climate is what you expect, weather is what you get*"

- Climate effects: spatial marginal (tail) distributions
- Weather effects: spatial (extremal) dependence strucure

Marginal and Dependence Modeling

- Marginal modeling: model the site–wise distribution of extremes by using extreme value distribution (i.e. GEV, GPD)

- Dependence modeling: model the spatial dependence of extreme values



Figures courtesy of Jenny Wadsworth

Marginal Modeling of Spatial Extremes

- Have observations at $\mathbf{s}_1, \cdots, \mathbf{s}_k \in \mathcal{S}$, can fit a GEV (or GPD) at each location separately. But won't be able say something for those unobserved locations

- Need a spatial model to characterize how the extreme value distribution (e.g. $\mathrm{GEV}(\mu(s), \sigma(s), \xi(s))$) varies in space

- Sensible marginal modeling allows for

  1. spatial interpolation of marginal distributions (e.g. return level map)

  2. borrowing strength to reduce estimation uncertainty

Outline

Spatial Trend Surface Approach

Formulate a regression function for each parameter:

$$\mu(s) = \beta_{0,\mu} + \sum_{i=1}^{p} x_i(s)\beta_{i,\mu}$$

$$\sigma(s) = \beta_{0,\sigma} + \sum_{j=1}^{q} x_j(s)\beta j, \sigma$$

$$\xi(s) = \beta_{0,\xi} + \sum_{k=1}^{r} x_k(s)\beta_{k,\xi}$$

- Need to observe the covariates $x's$ everywhere in the spatial domain. Lon/Lat/Alt are typically used
- Common modeling practice: $p > q > r$. Usually set $r = 0$ (i.e. constant shape parameter)

Bayesian Hierarchical Approach

- **Data**: $M(\mathbf{s}_i) | (\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})) \stackrel{ind}{\sim} \mathrm{GEV}(\mu(\mathbf{s}_i), \sigma(\mathbf{s}_i), \xi(\mathbf{s}_i))$, $\mathbf{s}_i \in \mathcal{S}$

- **Latent processes**: $(\mu(\mathbf{s}), \log(\sigma(\mathbf{s})))$ are modeled Gaussian Processes $\mathrm{GP}(m_{\boldsymbol{\beta}}(\cdot), K_{\boldsymbol{\theta}}(\cdot, \cdot))$ to account for "non-parametric" spatial regression functions

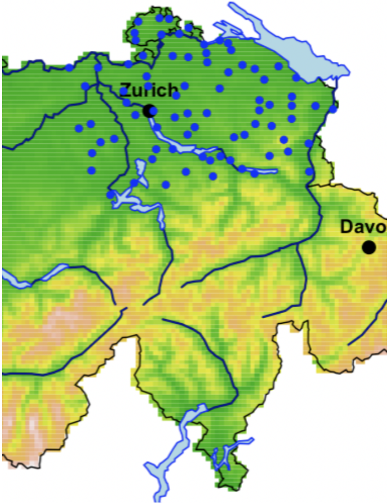
- **Prior**: need to specify the prior distributions for all parameters

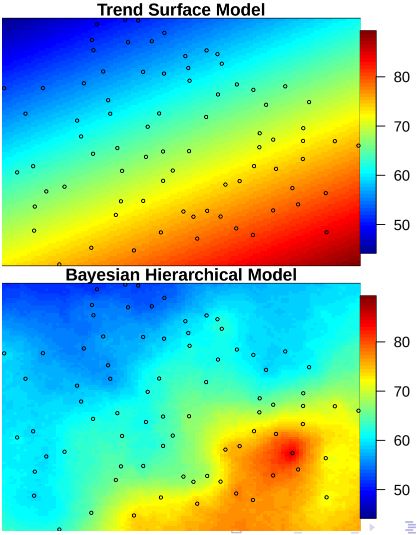
Use MCMC algorithm to carry out inference (e.g., [Cooley et al., 2007; Cooley and Sain, 2010])

Estimating Switzerland Rainfall Return Level Map

[Davison, Padoan & Ribatet, 2012]

Estimate 20-year return level map based on June–Aug. annual maxima precipitation over the years 1962–2008





Summary of Bayesian Hierarchical Approach

- $+$: The latent variable approach is very flexible and can helps in improving "weak" trend surfaces
- $+$: Easily extends to threshold exceedances
- -: No spatial dependence (on the data layer) is taken into account
- -: Quite difficult to define prior distributions and identifiability problems

Outline

Extremes of Spatial Processes

So far the (data level) spatial dependence in extremes has been ignored, we'd like to have some models to characterize spatial extremal dependence (e.g. spatial dependence for annual maxima)

- **Want**: a limiting stochastic process $\{M(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ s.t.

$$\{\frac{\max_{t=1}^{n} X_t(\boldsymbol{s}) - b_n(\boldsymbol{s})}{a_n(\boldsymbol{s})}\}_{\boldsymbol{s}\in\mathcal{S}} \stackrel{fdd}{\to} \{M(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$$

  $\Rightarrow$ if $\{M(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$ exists, it must be a max-stable process

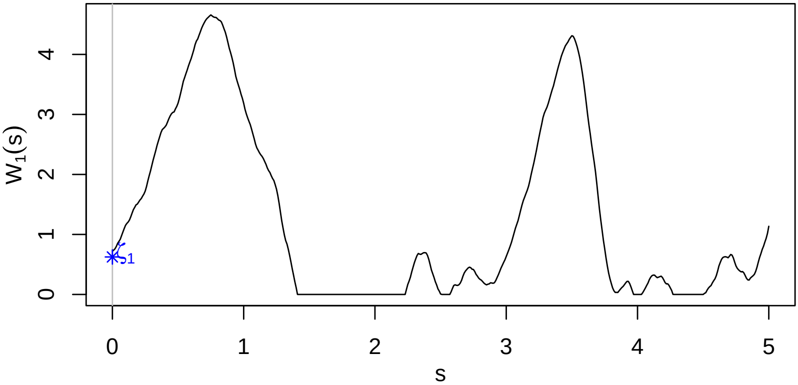
- $\{M(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$: pointwise maxima over the spatial domain $\Rightarrow$ $\approx \mathrm{GEV}$ spatial marginals, still need to model the spatial depedence strcuture

- Any max-stable process (MSP) is specified by its spectral representation $\Rightarrow$ it is easier to construct a max-stable process model via the spectral representation

Spectral Representations for MSP [de Haan, 1984; Schlather, 2002]

$$\left\{ \max_i \zeta_i W_i(\boldsymbol{s}) \right\}_{\boldsymbol{s} \in \mathbb{R}^k},$$

- ζ: Poisson process on $(0, \infty]$ with intensity ζ^{-2}
- $W(s)$: non-negative stochastic process s.t. $\mathbb{E}[W(\boldsymbol{s})] = 1 \, \forall \boldsymbol{s}$

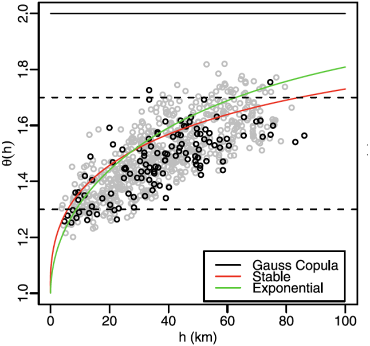


Exploring Spatial Dependence via Extremal Coefficient

The extremal coefficient is

$$\theta_d = \frac{\log \mathbb{P}(\max_{j=1}^{d} M(\boldsymbol{s}_j) \leq m)}{\log \mathbb{P}(M(\boldsymbol{s}) \leq m)} = V(1, \cdots, 1).$$

- 1 (totally dependent) $\leq \theta_d \leq d$ (independent)
- Can quantify spatial dependence by considering θ_2 vs. distance



- We expect θ_2 to rise from 1 to 2 as space distance increases
- $\chi_h = 2 - \theta_2(h)$, $\chi_h = \lim_{u \to 1} \chi_h(u) = \mathbb{P}\left(F_M(\boldsymbol{s}_i) > u | F_M(\boldsymbol{s}_j) > u\right)$, where $h = |\boldsymbol{s}_i - \boldsymbol{s}_j|$

Source: Davison et al. 2012

Parameter Estimation: Likelihood for a Max-Stable Process

Joint cdf:

$$\mathbb{P}(M_1 \le m_1, \cdots .M_d \le m_d) = \exp(-\underbrace{\mathbb{E}\left[\max\left\{\frac{W(\boldsymbol{s}_1)}{m_1}, \cdots, \frac{W(\boldsymbol{s}_d)}{m_d}\right\}\right]}_{V(m_1,\cdots,m_d)\equiv V})$$

$\Rightarrow$ joint density

$$\frac{\partial}{\partial m_1 \cdots \partial m_d} e^{-V} = e^{-V} \sum_{\pi \in \mathcal{P}_d} (-1)^{|\tau|} \prod_{j=1}^{|\tau|} \frac{\partial^{|\tau_j|}}{\partial m_{\tau_j}} V$$

where the cardinality of $\mathcal{P}_d = B_d$, the Bell number of order d

$$\begin{aligned} B_3 &= 5 \\ B_{10} &= 115,975! \\ B_{25} &= 4,638,590,332,229,999,353!!! \end{aligned}$$

Current Modeling Practice on Spatial Extremes

- Composite likelihood (Lindsay, 1988): pairwise likelihoods (Padoan et al. (2010)), Triplewise likelihoods (Genton et al. (2011), Huser and Davison (2013))
- Bayesian hierarchical approach for max-stable processes: Reich and Shaby (2012)
- Full likelihood by exploiting “temporal” information: Wadsworth and Tawn (2014), Thibaud et al. (2015)

Fitting Max-Stable Porcesses using Pairwise Likelihood

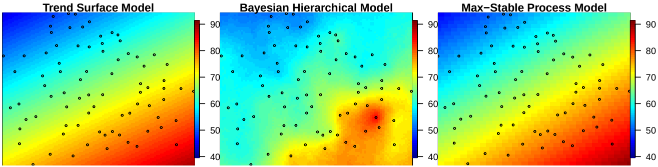
Since the (full) likelihood is too expensive to compute, one can consider pairwise log-likelihood

$$\ell_p(\theta; \boldsymbol{m}) = \sum_{t=1}^{T} \sum_{j=1}^{d-1} \sum_{k=j+1}^{d} w_{jk} \log g(m_t^{(j)}, m_t^{(k)}; \theta)$$

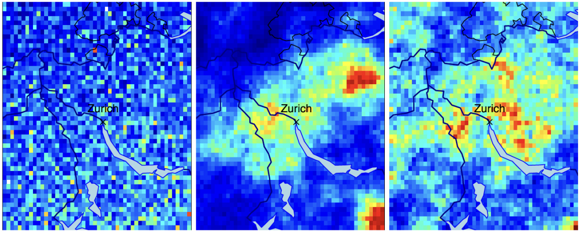
- g is the density of the associated bivariate GEV where spatial covariates can be added to define trend surfaces
- Maximize this gives the maximum composite likelihood estimator (MCLE)
- Inference can be performed under the composite likelihood framework

"Climate" and "Weather" Spatial Predictions

Climate: 20-year return level estimates



Weather: One realization of annual maximum field from BHM and two other MSP models



Source: Davioson et. al, 2012

Summary of Max-Stable Process Models

- $+$: Justified by extreme value theory
- $+$: Able to describe asymptotic dependence
- -: Describe everything that is asymptotically independent as exactly independent
- -: Full likelihood inference not feasible for "large" spatial data $\Rightarrow$ pairwise likelihood used instead

Summary of Spatial Extremes

- "Climate" vs. "Weather" Spatial Effects
- Bayesian hierarchical models and max-stable processes can be used to deal with climate and weather spatial effects, respectively
- Computation can be challenging: MCMC required for fitting Bayesian hierarchical models; while composite likelihood approach is needed for fitting MSP

Outline

Takeaway Message

- An extreme value analysis (typically) uses only data considered to be extreme $\Rightarrow$ large uncertainties are intrinsic to the problem
- Distributions for tail modeling are justified by asymptotic results from extreme value theory
- Return level is typically used for communicating risk
- Tail dependence is described very differently than dependence in the central part of the distribution
- Spatial models (for both climate and weather problem) can be fitted (with some effort) to high-dimensional problems

Things Not Address

- Point process representation
- Bayesian inference
- Unconditional/conditional simulation of MSP
- Spatial models for handling both AD and AI

References I

Books/Book Chapters

- Coles (2001) An Introduction to Statistical Modelling of Extreme Values. Springer
- Beirlant, Goegebeur, Segers & Teugels (2004). Statistics of extremes: theory and applications. John Wiley & Sons
- de Haan & Ferreira (2006) Extreme Value Theory. Springer
- Cooley, Hunter, & Smith (2019). Univariate and Multivariate Extremes for the Environmental Sciences. Handbook of environmental and ecological statistics Chapman and Hall/CRC
- Davison, Huser, & Thibaud (2019). Spatial extremes. Handbook of environmental and ecological statistics Chapman and Hall/CRC

Review Papers

- Davison et. al (2012) Stat. Sci.; Cooley et. al (2012) REVSTAT
- Davison & Huser (2015) Annu. Rev. Stat. Appl.

References II

Univariate Extremes

- Fisher & Tippet (1928) Math. Proc. Cambridge Philos. Soc.
- Gnedenko (1943) Ann. of Math.
- Gumbel (1958) Statistics of Extremes. Columbia Univ. Press
- Davison & Smith (1990) JRSSB

Multivariate Extremes

- Barnett (1976) JRSSA
- Tawn (1988) Biometrika
- Heffernan & Tawn (2004) JRSSB

Spatial Extremes

- Cooley, Nychka & Naveau, (2007) JASA
- Cooley & Sain (2010) JABES
- de Haan (1984) Ann. Probab; Schlather (2002) Extremes
- Padoan, Ribatet & Sisson (2010) JASA