Statistical Methods for Analyzing Extremes

Instructor: Whitney Huang

School of MATHEMATICAL AND STATISTICAL SCIENCES Clemson University

✉: wkhuang@clemson.edu

🖱: https://whitneyhuang83.github.io/

National Science Foundation WHERE DISCOVERIES BEGIN

Collaborative Strategies for Predicting and Measuring Uncertainty in Rare Occurrences in Civil and Environmental Systems

Golden, CO COLORADO SCHOOL OF MINES, November 6 2024

STATMOS samsi NSF Duke NCSU UNC NCAR PACIFIC CLIMATE IMPACTS CONSORTIUM University of Victoria CANSSI INCASS

Agenda

Outline

Introduction to Statistical Modeling

Goals: To learn from data by characterizing, predicting, and simulating processes of interest.

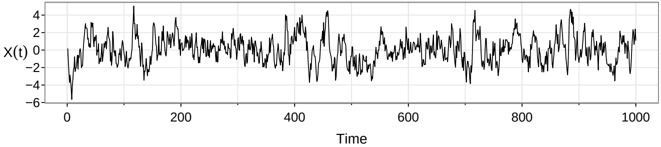
Purposes:

- To provide a comprehensive description of **data**
- To quantify the uncertainty in inferring the **process**
- To overcome the lack of data (prediction, data fusion, extrapolation)
- To complement physics-based models (also known as UQ)

Approach: To build **statistical models**, which are mathematical representations that describe how **data** is generated from the underlying **process**.

Statistical Modeling Workflow

1. **Problem formulation and exploratory data analysis**

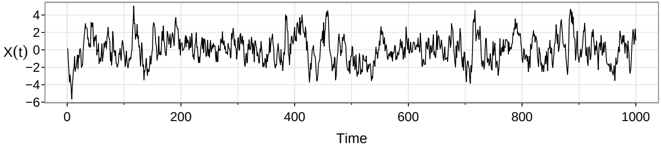


2. **Construction of a statistical model:**
   Example: AR(1) $X_{t+1} = \phi X_t + \sigma \epsilon_{t+1}, t \geq 0$
3. **Estimation of the model parameters** (ϕ, σ^2)**:** Least square, Maximum likelihood, Bayesian Methods,...
4. **Evaluation of fitted model:** QQ plot, out-of-sample validation, mean-squared error, scoring rules, ...
5. **Inference, prediction, simulation** AR(1):
   - Prediction: $\hat{x}_{t+1} = \hat{\phi} x_t$
   - Simulation: $x_0 \sim P_0$, $\epsilon \sim P_\epsilon$, for (t in $1:n$) $x_{t+1} = \hat{\phi} x_t + \hat{\sigma}\epsilon_{t+1}$

Let's focus on **extreme value analysis** for the rest of this course

Statistical Modeling Workflow

1. **Problem formulation and exploratory data analysis**

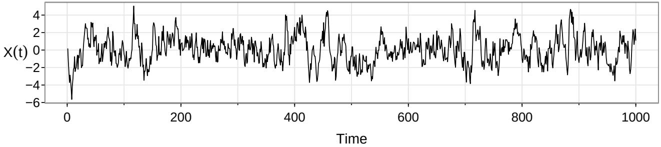


2. **Construction of a statistical model:**
   Example: AR(1) $X_{t+1} = \phi X_t + \sigma \epsilon_{t+1}, t \geq 0$
3. **Estimation of the model parameters** (ϕ, σ^2)**:** Least square, Maximum likelihood, Bayesian Methods,...
4. **Evaluation of fitted model:** QQ plot, out-of-sample validation, mean-squared error, scoring rules, ...
5. **Inference, prediction, simulation** AR(1):
   - Prediction: $\hat{x}_{t+1} = \hat{\phi} x_t$
   - Simulation: $x_0 \sim P_0$, $\epsilon \sim P_\epsilon$, for (t in $1 : n$) $x_{t+1} = \hat{\phi} x_t + \hat{\sigma} \epsilon_{t+1}$

Let's focus on **extreme value analysis** for the rest of this course

Statistical Modeling Workflow

1. **Problem formulation and exploratory data analysis**

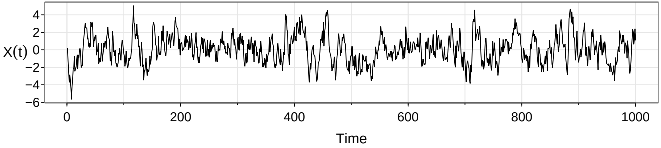


2. **Construction of a statistical model:**
   Example: AR(1) $X_{t+1} = \phi X_t + \sigma \epsilon_{t+1}, t \geq 0$
3. **Estimation of the model parameters (ϕ, σ^2):** Least square, Maximum likelihood, Bayesian Methods,...
4. **Evaluation of fitted model:** QQ plot, out-of-sample validation, mean-squared error, scoring rules, ...
5. **Inference, prediction, simulation** AR(1):
   - Prediction: $\hat{x}_{t+1} = \hat{\phi} x_t$
   - Simulation: $x_0 \sim P_0$, $\epsilon \sim P_\epsilon$, for (t in 1 : n) $x_{t+1} = \hat{\phi} x_t + \hat{\sigma} \epsilon_{t+1}$

Let's focus on **extreme value analysis** for the rest of this course

Statistical Modeling Workflow

1. **Problem formulation and exploratory data analysis**

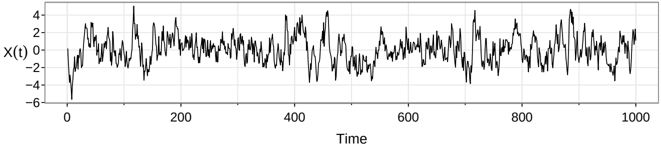


2. **Construction of a statistical model:**
   Example: AR(1) $X_{t+1} = \phi X_t + \sigma \epsilon_{t+1}, t \geq 0$
3. **Estimation of the model parameters** (ϕ, σ^2)**:** Least square, Maximum likelihood, Bayesian Methods,...
4. **Evaluation of fitted model:** QQ plot, out-of-sample validation, mean-squared error, scoring rules, ...
5. **Inference, prediction, simulation** AR(1):
   - Prediction: $\hat{x}_{t+1} = \hat{\phi} x_t$
   - Simulation: $x_0 \sim P_0$, $\epsilon \sim P_\epsilon$, for (t in 1 : n) $x_{t+1} = \hat{\phi} x_t + \hat{\sigma} \epsilon_{t+1}$

Let's focus on **extreme value analysis** for the rest of this course

Statistical Modeling Workflow

1. **Problem formulation and exploratory data analysis**

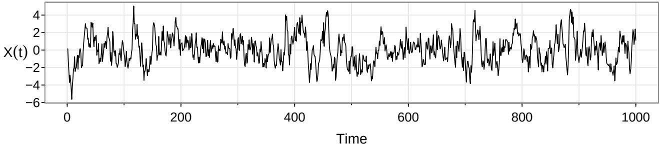


2. **Construction of a statistical model:**
   Example: AR(1) $X_{t+1} = \phi X_t + \sigma \epsilon_{t+1}, t \geq 0$
3. **Estimation of the model parameters** (ϕ, σ^2)**:** Least square, Maximum likelihood, Bayesian Methods,...
4. **Evaluation of fitted model:** QQ plot, out-of-sample validation, mean-squared error, scoring rules, ...
5. **Inference, prediction, simulation** AR(1):
   - Prediction: $\hat{x}_{t+1} = \hat{\phi} x_t$
   - Simulation: $x_0 \sim P_0$, $\epsilon \sim P_\epsilon$, for (t in $1:n$) $x_{t+1} = \hat{\phi} x_t + \hat{\sigma} \epsilon_{t+1}$

Let's focus on **extreme value analysis** for the rest of this course

Statistical Modeling Workflow

1. **Problem formulation and exploratory data analysis**

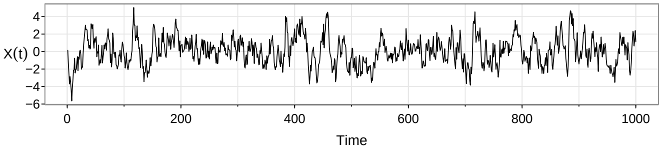


2. **Construction of a statistical model:**
   Example: AR(1) $X_{t+1} = \phi X_t + \sigma \epsilon_{t+1}, t \geq 0$
3. **Estimation of the model parameters** (ϕ, σ^2)**:** Least square, Maximum likelihood, Bayesian Methods,...
4. **Evaluation of fitted model:** QQ plot, out-of-sample validation, mean-squared error, scoring rules, ...
5. **Inference, prediction, simulation** AR(1):
   - Prediction: $\hat{x}_{t+1} = \hat{\phi} x_t$
   - Simulation: $x_0 \sim P_0$, $\epsilon \sim P_\epsilon$, for (t in $1:n$) $x_{t+1} = \hat{\phi} x_t + \hat{\sigma}\epsilon_{t+1}$

Let's focus on **extreme value analysis** for the rest of this course

Statistical Modeling Workflow

1. **Problem formulation and exploratory data analysis**

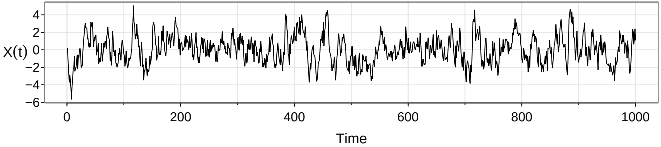


2. **Construction of a statistical model:**
   Example: AR(1) $X_{t+1} = \phi X_t + \sigma \epsilon_{t+1}, t \geq 0$
3. **Estimation of the model parameters** (ϕ, σ^2)**:** Least square, Maximum likelihood, Bayesian Methods,...
4. **Evaluation of fitted model:** QQ plot, out-of-sample validation, mean-squared error, scoring rules, ...
5. **Inference, prediction, simulation** AR(1):
   - Prediction: $\hat{x}_{t+1} = \hat{\phi} x_t$
   - Simulation: $x_0 \sim P_0$, $\epsilon \sim P_\epsilon$, for (t in $1:n$) $x_{t+1} = \hat{\phi} x_t + \hat{\sigma} \epsilon_{t+1}$

Let's focus on **extreme value analysis** for the rest of this course

Statistical Modeling Workflow

1. **Problem formulation and exploratory data analysis**

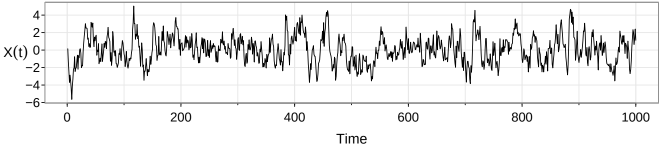


2. **Construction of a statistical model:**
   Example: AR(1) $X_{t+1} = \phi X_t + \sigma \epsilon_{t+1}, t \geq 0$
3. **Estimation of the model parameters** (ϕ, σ^2)**:** Least square, Maximum likelihood, Bayesian Methods,...
4. **Evaluation of fitted model:** QQ plot, out-of-sample validation, mean-squared error, scoring rules, ...
5. **Inference, prediction, simulation** AR(1):
   - Prediction: $\hat{x}_{t+1} = \hat{\phi} x_t$
   - Simulation: $x_0 \sim P_0$, $\epsilon \sim P_\epsilon$, for (t in $1 : n$) $x_{t+1} = \hat{\phi} x_t + \hat{\sigma} \epsilon_{t+1}$

Let's focus on **extreme value analysis** for the rest of this course

Statistical Modeling Workflow

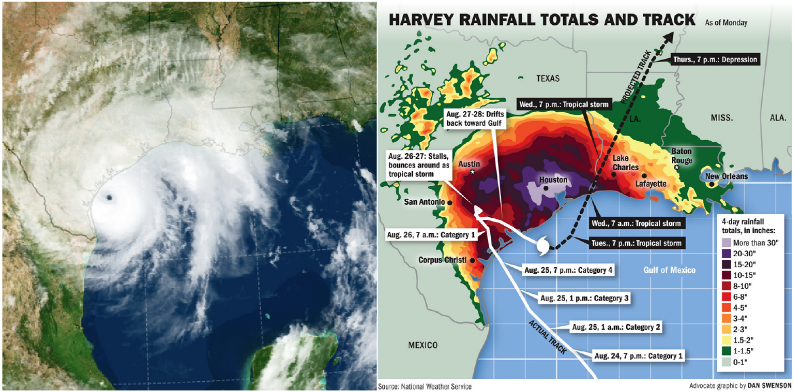
1. **Problem formulation and exploratory data analysis**



2. **Construction of a statistical model:**
   Example: AR(1) $X_{t+1} = \phi X_t + \sigma \epsilon_{t+1}, t \geq 0$
3. **Estimation of the model parameters** (ϕ, σ^2)**:** Least square, Maximum likelihood, Bayesian Methods,...
4. **Evaluation of fitted model:** QQ plot, out-of-sample validation, mean-squared error, scoring rules, ...
5. **Inference, prediction, simulation** AR(1):
   - Prediction: $\hat{x}_{t+1} = \hat{\phi} x_t$
   - Simulation: $x_0 \sim P_0$, $\epsilon \sim P_\epsilon$, for (t in $1:n$) $x_{t+1} = \hat{\phi} x_t + \hat{\sigma} \epsilon_{t+1}$

Let's focus on **extreme value analysis** for the rest of this course

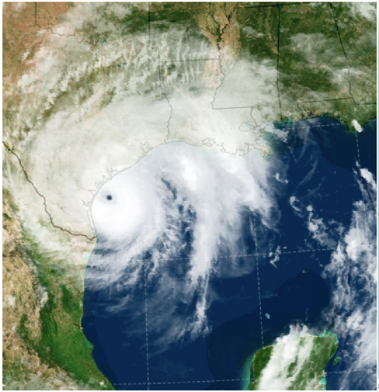
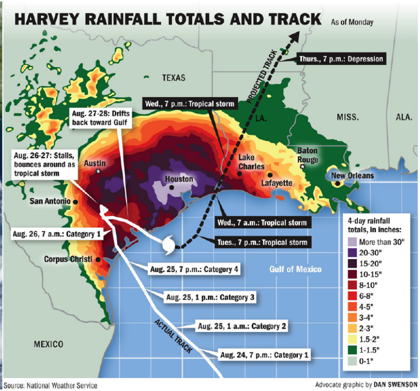
Extreme Rainfall During Hurricane Harvey



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

- The highest total rainfall was 60.6 inches (1539.7 mm) near Nederland, TX
- **Annual average** rainfall for Golden, CO: 18.7 inches (475 mm)

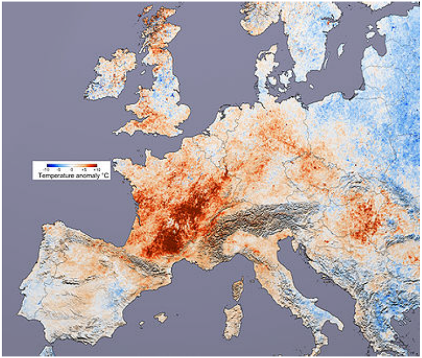
Extreme Rainfall During Hurricane Harvey

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

- The highest total rainfall was 60.6 inches (1539.7 mm) near Nederland, TX
- **Annual average** rainfall for Golden, CO: 18.7 inches (475 mm)

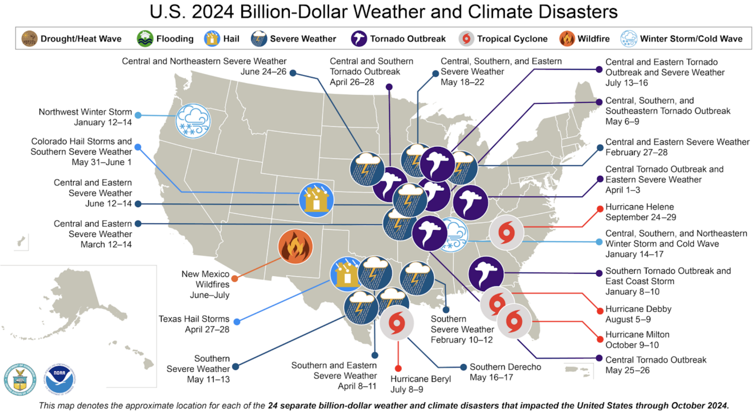
Some Examples of Weather/Climate Extremes



- **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.



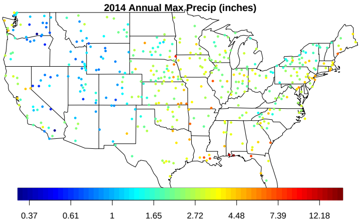
Source: National Oceanic and Atmospheric Administration

Goal: Quantify the tail behavior ⇒ may require extrapolation

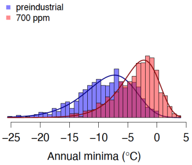
Some Scientific Questions

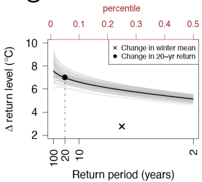
- How to estimate the magnitude of extreme events (e.g. 100-year rainfall)?

- How extremes vary in space?



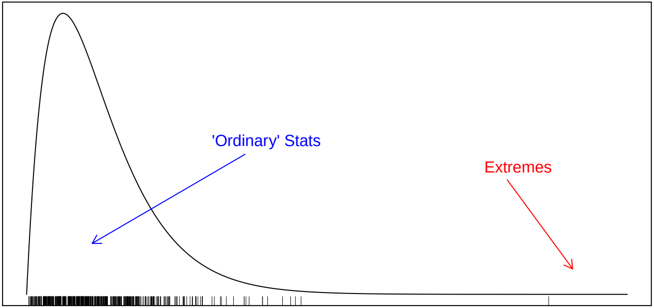
- How extremes change in future climate conditions?





Outline

Usual vs Extremes



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

Let's examine the distribution of the **sample maximum** to learn about extremes

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

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

$\Rightarrow$ 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

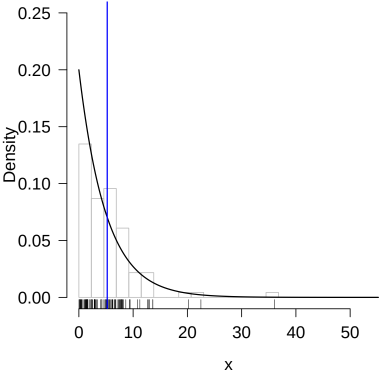
Question: Can we get the limiting distribution of

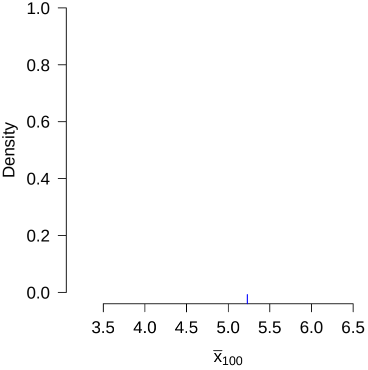
$$\frac{M_n - b_n}{a_n}$$

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]

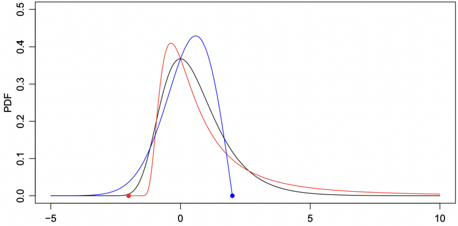
Define $M_n = \max\{X_1, \cdots, X_n\}$ where $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} F$. If $\exists\, a_n > 0$ and $b_n \in \mathbb{R}$ such that, as $n \to \infty$, if

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

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

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

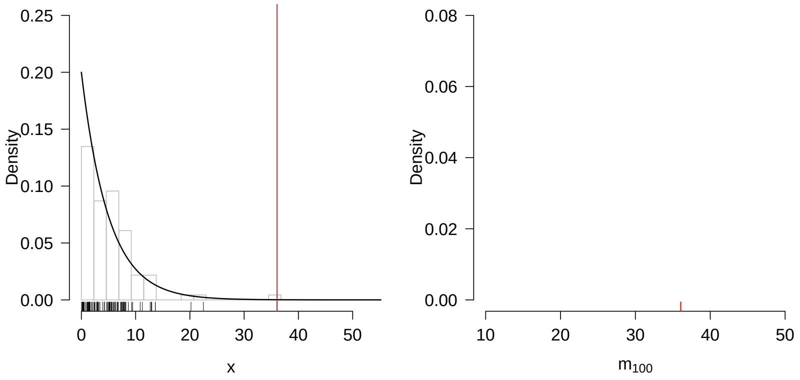
where $x_+ = \max(x, 0)$ and $G(x)$ is the distribution function of the generalized extreme value distribution ($\mathtt{GEV}(\mu, \sigma, \xi)$), where μ and σ are location and scale parameters, and ξ is the shape parameter



- $\xi > 0$: 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

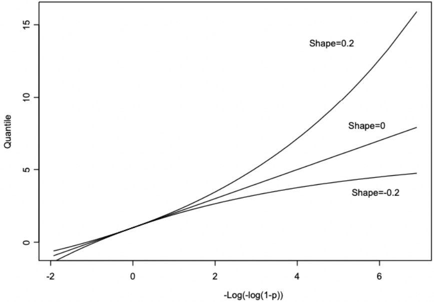


Quantiles and Return Levels

- 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: Modeling Block Maxima as GEV

Assume n is large enough so that

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

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

Uncertainty Quantification for GEV Estimation

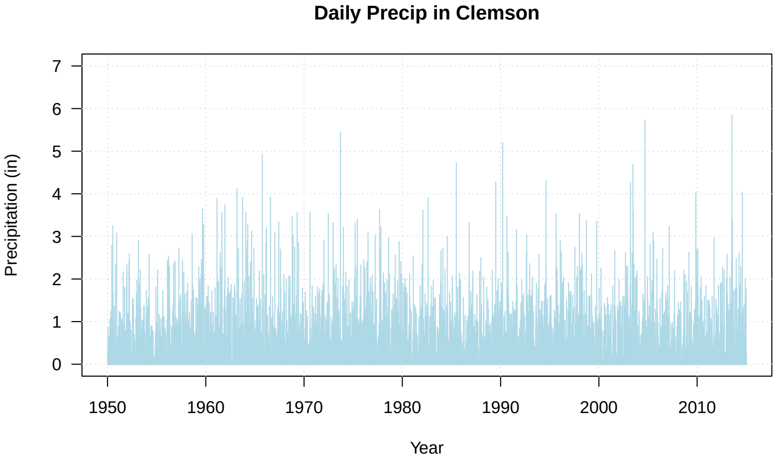
Parameters not very interpretable. Better to provide uncertainty about a more meaningful quantity (e.g. 100-year return level)

Two methods for constructing confidence intervals (CIs):

- **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[2]]

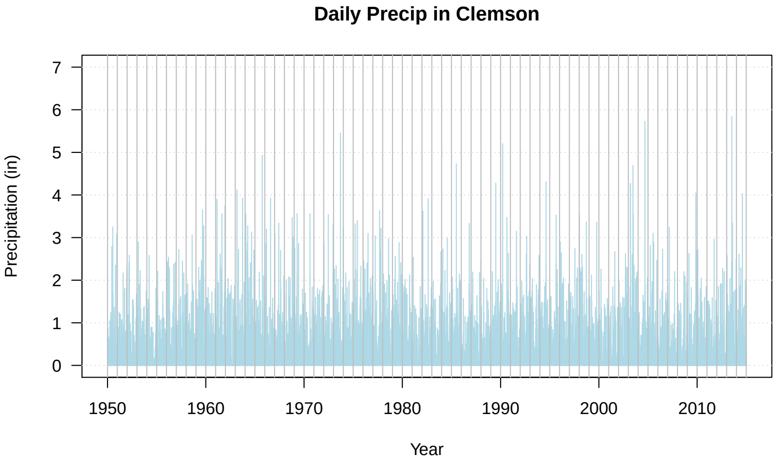
These data were compiled by Will Kleiber, thanks Will!



[2]United States Historical Climatology Network (http://cdiac.ornl.gov/epubs/ndp/ushcn/ushcn.html)

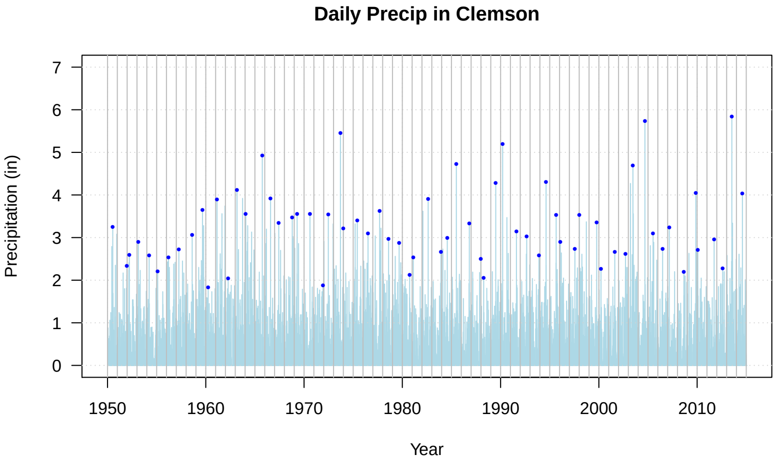
Block-Maximum Method [Gumbel 1958]

1. Determine the block size and extract the block maxima



Block-Maximum Method [Gumbel 1958]

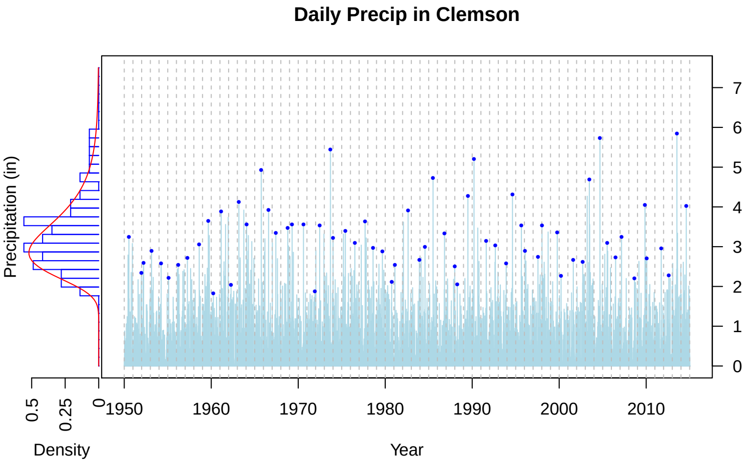
1. Determine the block size and extract the block maxima



Block-Maximum Method [Gumbel 1958]

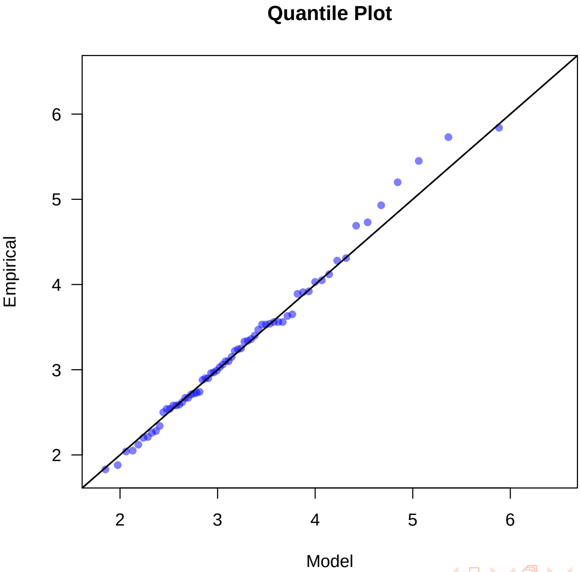
2. Fit the GEV to the maxima and assess the fit.

$$\hat{\mu} = 2.85(0.10), \quad \hat{\sigma} = 0.71(0.07), \quad \hat{\xi} = 0.01(0.10)$$



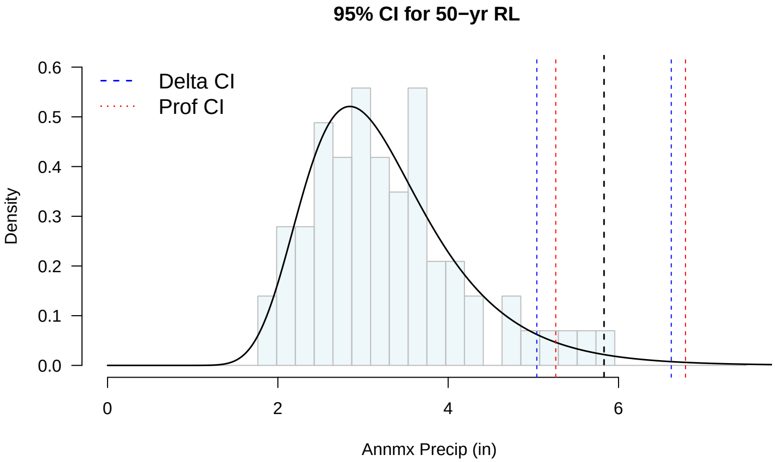
Block-Maximum Method [Gumbel 1958]

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

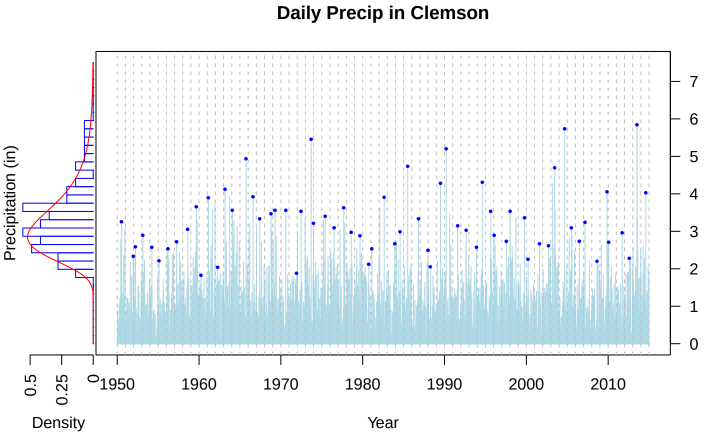


Block-Maximum Method [Gumbel 1958]

3. Perform inference for return levels



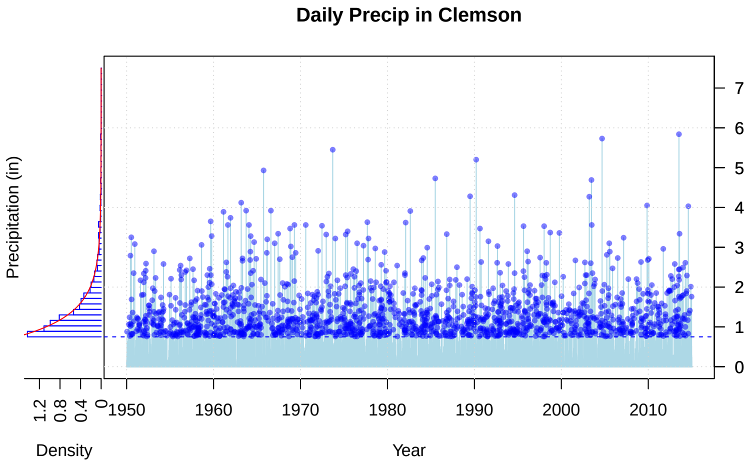
Can We Obtain More Data for Extreme Value Analysis?



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

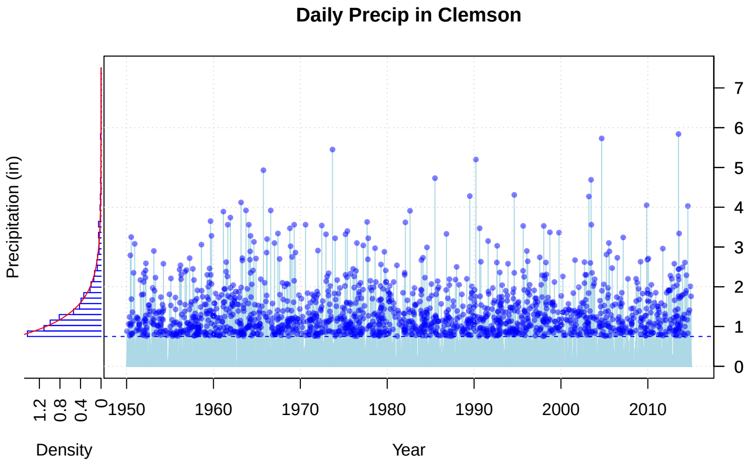
Threshold-Exceedance Method [Davison & Smith 1990]

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



Threshold-Exceedance Method [Davison & Smith 1990]

2. Fit an appropriate model to exceedances



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

If $M_n = \max_{1 \leq i \leq n}\{X_i\} \approx$ GEV(μ, σ, ξ), 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

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

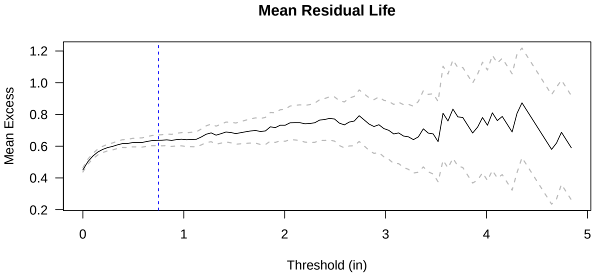
and $\tilde{\sigma} = \sigma + \xi(u - \mu)$

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

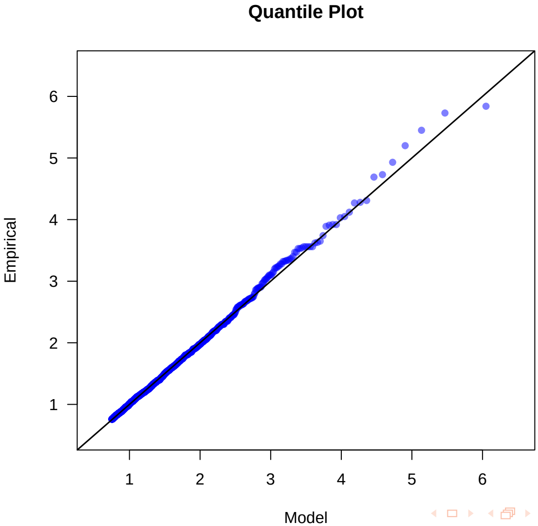


Task: To choose a u_0 s.t. the **mean residual life curve** starts to behave linearly for all $u > u_0$

Threshold-Exceedance Method [Davison & Smith 1990]

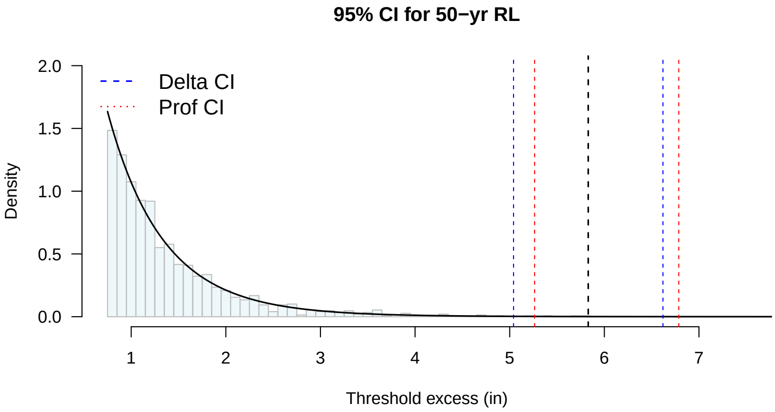
2. Fit the GPD to exceedances and assess the fit

$$\hat{\tilde{\sigma}} = 0.61(0.02), \quad \hat{\xi} = 0.05(0.03)$$



Threshold-Exceedance Method [Davison & Smith 1990]

3. Perform inference for return levels



Outline

Dependence and Non-Stationary

Methods introduced so far assume the process is independent and identically distributed. However, environmental time series often deviate from these assumptions, commonly as:

- **Local temporal dependence**: Successive values in the time series are related, while distant values remain independent.
- **Long-term trends**: The underlying distribution gradually changes over time.
- **Seasonal variation**: The underlying distribution changes periodically over time.

These departures can be addressed by extending both theory and modeling

Modeling Extremes under (Local) Temporal Dependence

Suppose the process of interest is stationary (but not independent) with short-range temporal dependence, where mixing conditions hold (Leadbetter et al., 1983):

Block maximum approach	Threshold exceedance approach
GEV still correct for the marginal. Since block maximum data likely have negligible dependence, proceed as usual ☺	GPD is "correct" for the marginal. But extremes likely occur in clusters, affecting inference as likelihood assumes independence of threshold exceedances ☹

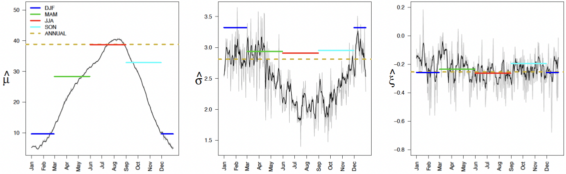
Modeling Threshold Exceedances

- **Declustering**: Filtering out an (approximately) independent set of threshold exceedances
- **Dependence Modeling**: Explicitly modeling the temporal dependence in the process

Nonstationary Extremes: Seasonality and Trend

Extreme value regression models are typically employed

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



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

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

  and let $\xi(t)$ be a constant

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

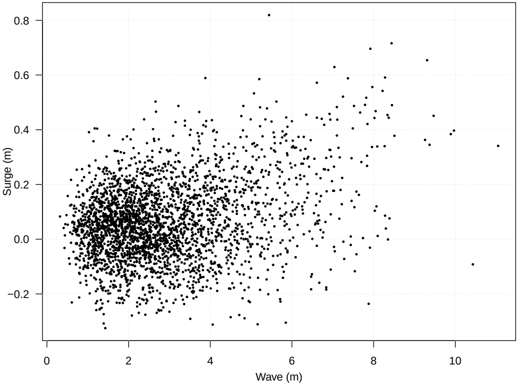
Remarks on Univariate Extremes

- To estimate the tail, EVT uses only extreme observations $\Rightarrow$ block maxima and threshold exceedances[3]
- The shape parameter ξ is extremely important (especially for extrapolation) but is difficult to reliably estimate
- The **threshold exceedance approach** allows the user to retain more data than the **block-maximum approach**:
  - ☺ reducing the uncertainty of parameter estimates
  - ☹ modeling is more complex with threshold selection and handling temporal dependence

[3] But there have been some recent developments on bridging bulk and tail distributions.

Outline

Wave–Surge Data at Newlyn, UK

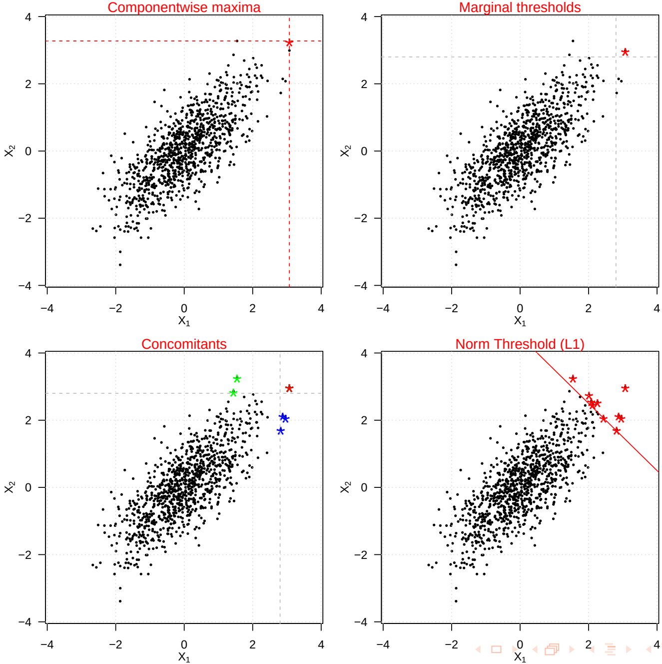


- Scientific motivation: many extremal problems are intrinsically multivariate

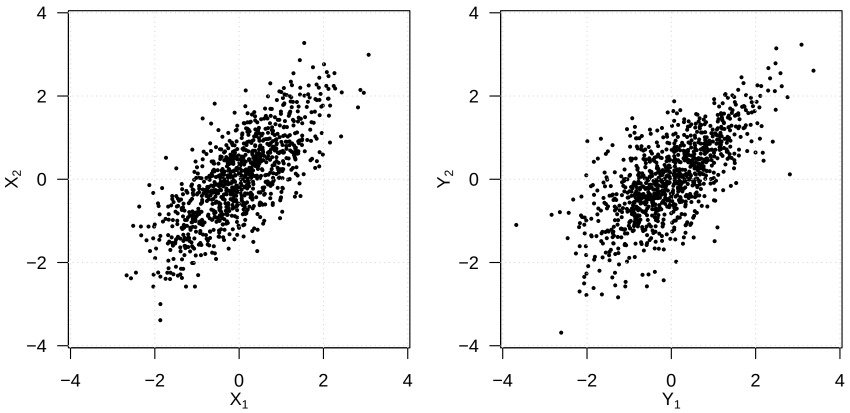
- Statistical motivation: uncertainty may be reduced by combining information from several sources

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

What is a Multivariate Extreme?

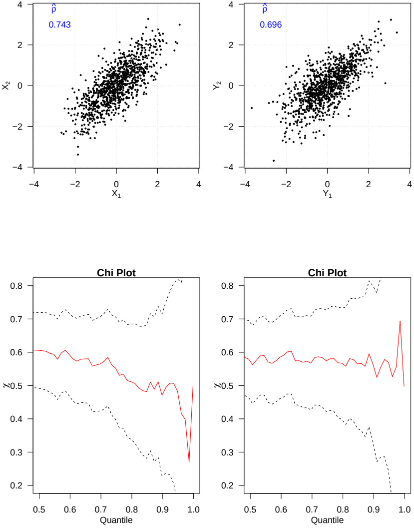


Describing Tail Dependence



A central aim of multivariate extremes is to describe tail dependence

Correlation Coefficient and Tail Dependence Parameter



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

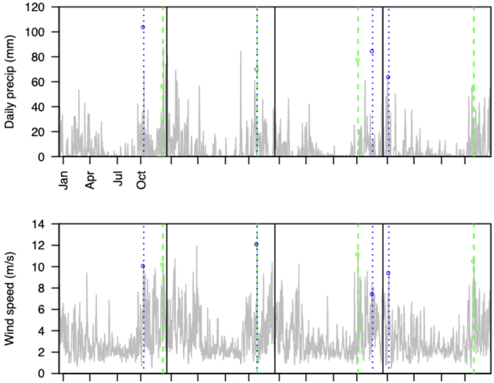
Upper tail dependence parameter:

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

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 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

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

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

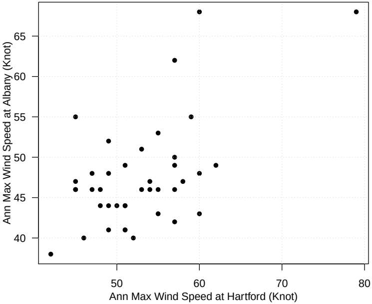
The Logistic Model

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

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

Annual Maximum Wind Speeds at Albany and Hartford

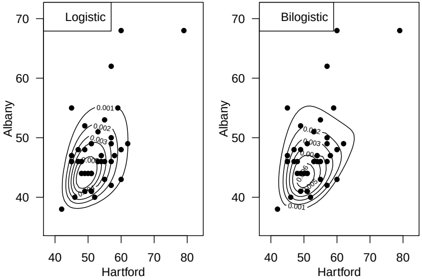


Marginal Parameter Estimates

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

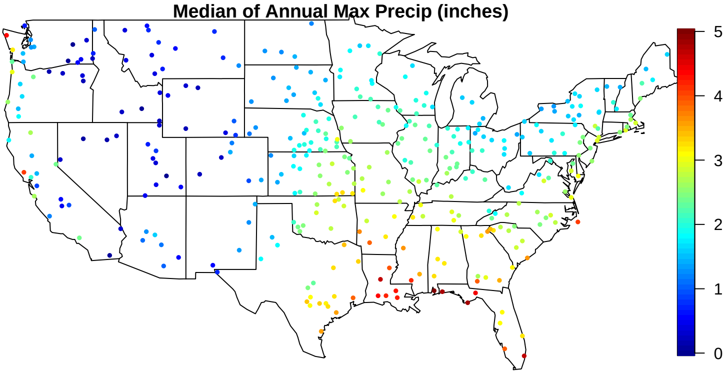


Remarks on Multivariate Extreme Value Analysis

- There is no unique definition of a multivariate extreme
- Tail dependence is NOT summarized with correlations; instead, we look at the upper tail dependence parameter χ
- Methodologies exist for both **block maxima** and **threshold exceedance approaches**

Outline

Climate and Weather Aspects of Spatial Extremes



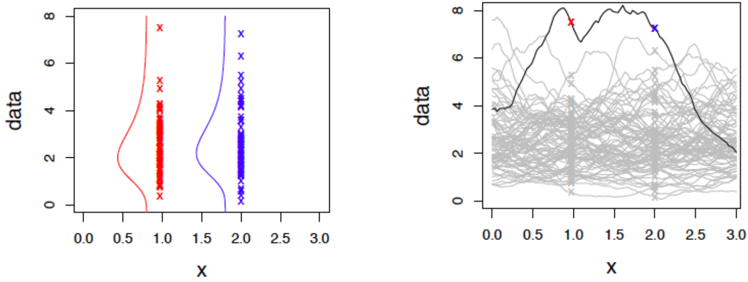
"*Climate is what you expect, weather is what you get*"

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

- **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

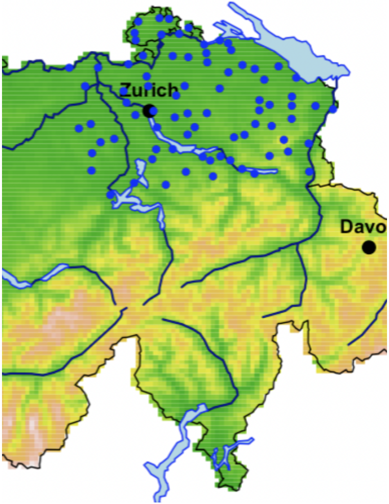
- **Prior**: need to specify the prior distributions for all parameters

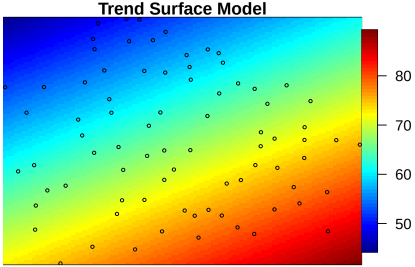
Use MCMC algorithm to carry out inference (e.g., [Cooley et al., 2007; Cooley and Sain, 2010])

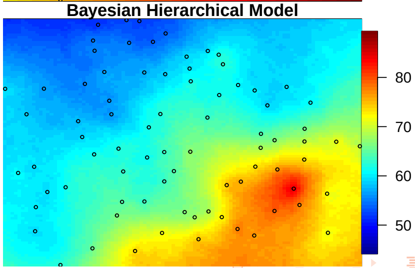
Estimating Switzerland Rainfall Return Level Map

[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

Summary of Spatial Extremes

- "Climate" (Marginal) vs. "Weather" (Dependence) spatial effects
- Bayesian hierarchical models can be used to deal with climate spatial effects
- MCMC is needed for fitting Bayesian hierarchical models, and computation can be challenging for big data problems

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 can be fitted, with some effort, to high-dimensional problems

Things Not Address

- Point process representation

  Underlying framework that unifies block maximum and threshold-exceedance methods

- Other approaches for multivariate extremes

- Theory and inference of max-stable processes

  But check a fairly recent opinion piece, "Modeling of spatial extremes in environmental data science: Time to move away from max-stable processes" by Huser et al.

- Spatial depdence models for handling both AD and AI

  Some experts in the house!

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